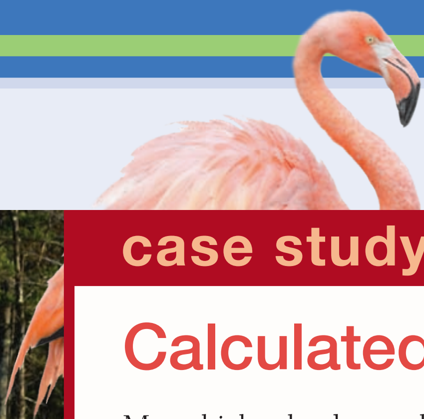
Probability: What Are the Chances?



case study

Calculated Risks

Many high schools now have drug-testing programs for athletes. The main goal of these programs is to reduce the use of banned substances by students who play sports. It is not practical to test every athlete for drug use regularly. Instead, school administrators give drug tests to randomly selected student athletes at unannounced times during the school year. Students who test positive face serious consequences, including letters to their parents, required counseling, and suspension from athletic participation.

Drug tests aren't perfect. Sometimes the tests say that athletes took a banned substance when they did not. This is known as a *false positive*. Other times, drug tests say that athletes are "clean" when they did take a banned substance. This is called a *false negative*.

Suppose that 16% of the high school athletes in a large school district have taken a banned substance. The drug test used by this district has a false positive rate of 5% and a false negative rate of 10%. If a randomly chosen athlete tests positive, what's the chance that the student actually took a banned substance?

By the end of the chapter, you will be ready to answer questions that involve drug and disease testing, false positives, and false negatives.

Introduction

Chance is all around us. You and your friend play rock-paper-scissors to determine who gets the last slice of pizza. A coin toss decides which team gets to receive the ball first in a football game. Many adults regularly play the lottery, hoping to win a big jackpot with a few lucky numbers. Others head to casinos or racetracks, hoping that some combination of luck and skill will pay off. People young and old play games of chance involving cards or dice or spinners. The traits that children inherit—gender, hair and eye color, blood type, handedness, dimples, whether they can roll their tongues—are determined by the chance involved in which genes get passed along by their parents.

A roll of a die, a simple random sample, and even genetic inheritance represent chance behavior that we can understand and work with. We can roll the die again and again and again. The outcomes are governed by chance, but in many repetitions a pattern emerges. We use mathematics to understand the regular patterns of chance behavior when we repeat the same chance process again and again.

The mathematics of chance is called *probability*. Probability is the topic of this chapter. Here is an Activity that gives you some idea of what lies ahead.

ACTIVITY The "1 in 6 wins" game

MATERIALS:
One six-sided die for each student

As a special promotion for its 20-ounce bottles of soda, a soft drink company printed a message on the inside of each bottle cap. Some of the caps said, "Please try again!" while others said, "You're a winner!" The company advertised the promotion with the slogan "1 in 6 wins a prize." The prize is a free 20-ounce bottle of soda, which comes out of the store owner's profits.

Seven friends each buy one 20-ounce bottle at a local convenience store. The store clerk is surprised when three of them win a prize. The store owner is concerned about losing money from giving away too many free sodas. She wonders if this group of friends is just lucky or if the company's 1-in-6 claim is inaccurate. In this Activity, you and your classmates will perform a *simulation* to help answer this question.

For now, let's assume that the company is telling the truth, and that every 20-ounce bottle of soda it fills has a 1-in-6 chance of getting a cap that says, "You're a winner!" We can model the status of an individual bottle with a six-sided die: let 1 through 5 represent "Please try again!" and 6 represent "You're a winner!"

1. Roll your die seven times to imitate the process of the seven friends buying their sodas. How many of them won a prize?

2. Your teacher will draw and label axes for a class dotplot. Plot the number of prize winners you got in Step 1 on the graph.

3. Repeat Steps 1 and 2 if needed to get a total of at least 40 repetitions of the simulation for your class.

4. Discuss the results with your classmates. What percent of the time did the friends come away with three or more prizes, just by chance? Does it seem plausible that the company is telling the truth, but that the seven friends just got lucky? Explain.

As the Activity shows, simulation is a powerful method for modeling chance behavior. Section 5.1 begins by examining the idea of probability and then illustrates how simulation can be used to estimate probabilities. In Sections 5.2 and 5.3, we develop the basic rules of probability. Along the way, we introduce some helpful tools for displaying possible outcomes from a chance process: two-way tables, Venn diagrams, and tree diagrams.

Probability calculations are the basis for inference. When we produce data by random sampling or randomized comparative experiments, the laws of probability answer the question "What would happen if we repeated the random sampling or random assignment process many times?" Many of the examples, exercises, and activities in this chapter focus on the connection between probability and inference.

5.1 Randomness, Probability, and Simulation

WHAT YOU WILL LEARN By the end of the section, you should be able to:

- Interpret probability as a long-run relative frequency.
- Use simulation to model chance behavior.

Toss a coin 10 times. How likely are you to get a run of 3 or more consecutive heads or tails? An airline knows that a certain percent of customers who purchased tickets will not show up for a flight. If the airline overbooks a particular flight, what are the chances that they'll have enough seats for the passengers who show up? A couple plans to have children until they have at least one child of each gender. How many children should they expect to have? To answer these questions, you need a better understanding of how chance behavior operates.

The Idea of Probability

In football, a coin toss helps determine which team gets the ball first. Why do the rules of football require a coin toss? Because tossing a coin seems a "fair" way to decide. That's one reason

why statisticians recommend random samples and randomized experiments. They avoid bias by letting chance decide who gets selected or who receives which treatment.

A big fact emerges when we watch coin tosses or the results of random sampling and random assignment closely: *chance behavior is unpredictable in the short run but has a regular and predictable pattern in the long run.* This remarkable fact is the basis for the idea of probability.

ACTIVITY | *Probability* applet

MATERIALS:
Computer with Internet connection

If you toss a fair coin, what's the chance that it shows heads? It's 1/2, or 0.5, right? In this Activity, you'll use the *Probability* applet at www.whfreeman.com/tps5e to investigate what probability really means.

1. If you toss a fair coin 10 times, how many heads will you get? Before you answer, launch the *Probability* applet. Set the number of tosses at 10 and click "Toss." What proportion of the tosses were heads? Click "Reset" and toss the coin 10 more times. What proportion of heads did you get this time? Repeat this process several more times. What do you notice?

2. What if you toss the coin 100 times? Reset the applet and have it do 100 tosses. Is the proportion of heads exactly equal to 0.5? Close to 0.5?

3. Keep on tossing without hitting "Reset." What happens to the proportion of heads?

4. As a class, discuss what the following statement means: "If you toss a fair coin, the probability of heads is 0.5."

5. Predict what will happen if you change the probability of heads to 0.3 (an unfair coin). Then use the applet to test your prediction.

6. If you toss a coin, it can land heads or tails. If you "toss" a thumbtack, it can land with the point sticking up or with the point down. Does that mean that the probability of a tossed thumbtack landing point up is 0.5? How could you find out? Discuss with your classmates.

We might suspect that a coin has probability 0.5 of coming up heads just because the coin has two sides. But we can't be sure. In fact, spinning a penny on a flat surface, rather than tossing the coin, gives heads a probability of about 0.45 rather than 0.5.[1] What about thumbtacks? They also have two ways to land—point up or point down—but the chance that a tossed thumbtack lands point up isn't 0.5. How do we know that? From tossing a thumbtack over and over and over again. Probability describes what happens in very many trials, and we must actually observe many tosses of a coin or thumbtack to pin down a probability.

EXAMPLE Tossing Coins

Short-run and long-run behavior

When you toss a coin, there are only two possible outcomes, heads or tails. Figure 5.1(a) shows the results of tossing a coin 20 times. For each number of tosses from 1 to 20, we have plotted the proportion of those tosses that gave a head. You can see that the proportion of heads starts at 1 on the first toss, falls to 0.5 when the second toss gives a tail, then rises to 0.67, and then falls to 0.5, and 0.4 as we get two more tails. After that, the proportion of heads continues to fluctuate but never exceeds 0.5 again.

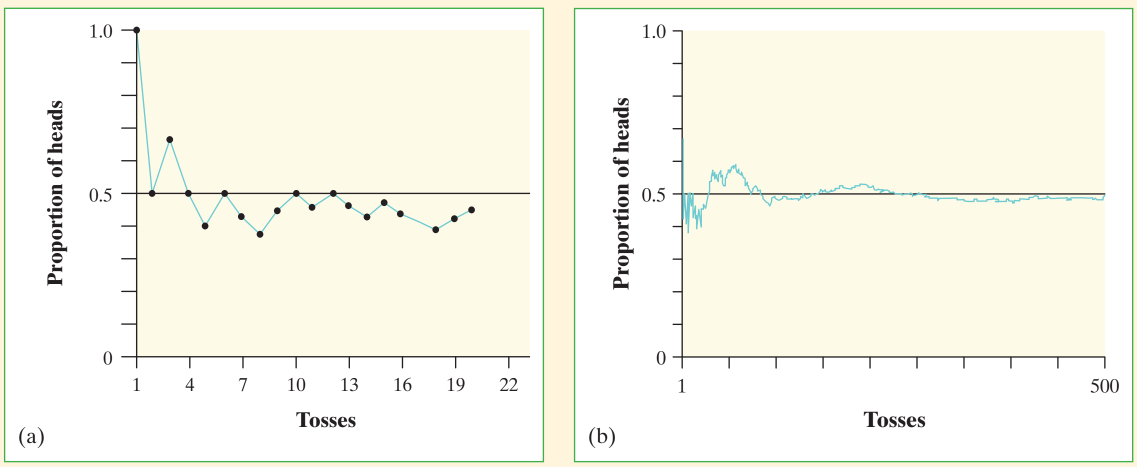


FIGURE 5.1 (a) The proportion of heads in the first 20 tosses of a coin. (b) The proportion of heads in the first 500 tosses of a coin.

Suppose we keep tossing the coin until we have made 500 tosses. Figure 5.1(b) shows the results. The proportion of tosses that produce heads is quite variable at first. As we make more and more tosses, however, the proportion of heads gets close to 0.5 and stays there.

The fact that the proportion of heads in many tosses eventually closes in on 0.5 is guaranteed by the **law of large numbers**. This important result says that if we observe more and more repetitions of any chance process, the proportion of times that a specific outcome occurs approaches a single value. We call this value the **probability**. The previous example confirms that the probability of getting a head when we toss a fair coin is 0.5. Probability 0.5 means "occurs half the time in a very large number of trials."

DEFINITION: Probability

The **probability** of any outcome of a chance process is a number between 0 and 1 that describes the proportion of times the outcome would occur in a very long series of repetitions.

Outcomes that never occur have probability 0. An outcome that happens on every repetition has probability 1. An outcome that happens half the time in a very long series of trials has probability 0.5. Of course, we can never observe a probability exactly. We could always continue tossing the coin, for example. The mathematics of probability is based on imagining what would happen in an indefinitely long series of trials.

Recall from Chapter 4 that "random" doesn't mean haphazard. In statistics, random means "by chance."

Probability gives us a language to describe the long-term regularity of random behavior. The outcome of a coin toss and the gender of the next baby born in a local hospital are both random. So is the result of a random sample or a random assignment. Even life insurance, for example, is based on the fact that deaths occur at random among many individuals.

EXAMPLE Life Insurance

Probability and risk

How do insurance companies decide how much to charge for life insurance? We can't predict whether a particular person will die in the next year. But the National Center for Health Statistics says that the proportion of men aged 20 to 24 years who die in any one year is 0.0015. This is the *probability* that a randomly selected young man will die next year. For women that age, the probability of death is about 0.0005. If an insurance company sells many policies to people aged 20 to 24, it knows that it will have to pay off next year on about 0.15% of the policies sold to men and on about 0.05% of the policies sold to women. Therefore, the company will charge about three times more to insure a man because the probability of having to pay is three times higher.

We often encounter the unpredictable side of randomness in our everyday lives, but we rarely see enough repetitions of the same chance process to observe the long-run regularity that probability describes. Life insurance companies, casinos, and others who make important decisions based on probability rely on the long-run predictability of random behavior.

CHECK YOUR UNDERSTANDING

1. According to the Book of Odds Web site www.bookofodds.com, the probability that a randomly selected U.S. adult usually eats breakfast is 0.61.

(a) Explain what probability 0.61 means in this setting.

(b) Why doesn't this probability say that if 100 U.S. adults are chosen at random, exactly 61 of them usually eat breakfast?

2. Probability is a measure of how likely an outcome is to occur. Match one of the probabilities that follow with each statement. Be prepared to defend your answer.

0 0.01 0.3 0.6 0.99 1

(a) This outcome is impossible. It can never occur.

(b) This outcome is certain. It will occur on every trial.

(c) This outcome is very unlikely, but it will occur once in a while in a long sequence of trials.

(d) This outcome will occur more often than not.

Myths about Randomness

The idea of probability seems straightforward. It answers the question "What would happen if we did this many times?" In fact, both the behavior of random phenomena and the idea of probability are a bit subtle.

ACTIVITY | Investigating randomness

1. Pretend that you are flipping a fair coin. Without actually flipping a coin, *imagine* the first toss. Write down the result you see in your mind, heads (H) or tails (T).
2. Imagine a second coin flip. Write down the result.
3. Keep doing this until you have recorded the results of 50 imaginary flips. Write your results in groups of 5 to make them easier to read, like this: HTHTH TTHHT, etc.
4. A *run* is a repetition of the same result. In the example in Step 3, there is a run of two tails followed by a run of two heads in the first 10 coin flips. Read through your 50 imagined coin flips and find the longest run.
5. Your teacher will draw and label a number line for a class dotplot. Plot the length of the longest run you got in Step 4 on the graph.
6. Use Table D, technology, or a coin to generate a similar list of 50 coin flips. Find the longest run that you have.
7. Your teacher will draw and label a number line with the same scale immediately above or below the one in Step 5. Plot the length of the longest run you got in Step 6 on the new dotplot.
8. Compare the distributions of longest run from imagined tosses and random tosses. What do you notice?

The idea of probability is that randomness is predictable in the long run. Unfortunately, our intuition about randomness tries to tell us that random phenomena should also be predictable in the short run. When they aren't, we look for some explanation other than chance variation.

EXAMPLE Runs in Coin Tossing

What looks random?

Toss a coin six times and record heads (H) or tails (T) on each toss. Which of the following outcomes is more probable?

HTHTTH TTTHHH

Almost everyone says that HTHTTH is more probable, because TTTHHH does not "look random." In fact, both are equally likely. That heads and tails are equally

probable says only that about half of a very long sequence of tosses will be heads. It doesn't say that heads and tails must come close to alternating in the short run. The coin has no memory. It doesn't know what past outcomes were, and it can't try to create a balanced sequence.

The outcome TTTHHH in tossing six coins looks unusual because of the runs of 3 straight tails and 3 straight heads. Runs seem "not random" to our intuition but are quite common. Here's a more striking example than tossing coins.

EXAMPLE That Shooter Seems "Hot"

Chance variation or skill?

Is there such a thing as a "hot hand" in basketball? Belief that runs must result from something other than "just chance" influences behavior. If a basketball player makes several consecutive shots, both the fans and her teammates believe that she has a "hot hand" and is more likely to make the next shot. Several studies have shown that runs of baskets made or missed are no more frequent in basketball than would be expected if the result of each shot is unrelated to the outcomes of the player's previous shots. If a player makes half her shots in the long run, her made shots and misses behave just like tosses of a coin—and that means that runs of makes and misses are more common than our intuition expects.[2]

Free throws may be a different story. A recent study suggests that players who shoot two free throws are *slightly* more likely to make the second shot if they make the first one.[3]

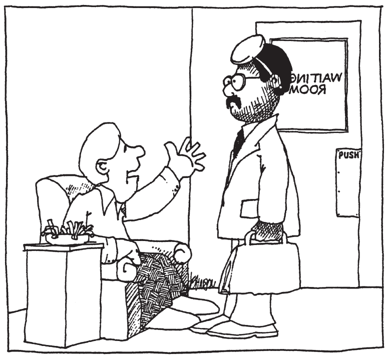
Once, at a convention in Las Vegas, one of the authors roamed the gambling floors, watching money disappear into the drop boxes under the tables. You can see some interesting human behavior in a casino. When the shooter in the dice game called craps rolls several winners in a row, some gamblers think she has a "hot hand" and bet that she will keep on winning. Others say that "the law of averages" means that she must now lose so that wins and losses will balance out. Believers in the law of averages think that if you toss a coin six times and get TTTTTT, the next toss must be more likely to give a head. It's true that in the long run heads will appear half the time. What is a myth is that future outcomes must make up for an imbalance like six straight tails.

Don't confuse the law of large numbers, which describes the big idea of probability, with the "law of averages" described here.

Coins and dice have no memories. A coin doesn't know that the first six outcomes were tails, and it can't try to get a head on the next toss to even things out. Of course, things do even out *in the long run*. That's the law of large numbers in action. After 10,000 tosses, the results of the first six tosses don't matter. They are overwhelmed by the results of the next 9994 tosses.

EXAMPLE Aren't We Due for a Boy?

Don't fall for the "law of averages"



"So the law of averages doesn't guarantee me a girl after seven straight boys, but can't I at least get a group discount on the delivery fee?"

Belief in this phony "law of averages" can lead to serious consequences. A few years ago, an advice columnist published a letter from a distraught mother of eight girls. She and her husband had planned to limit their family to four children, but they wanted to have at least one boy. When the first four children were all girls, they tried again—and again and again. After seven straight girls, even her doctor had assured her that "the law of averages was in our favor 100 to 1." Unfortunately for this couple, having children is like tossing coins. Eight girls in a row is highly unlikely, but once seven girls have been born, it is not at all unlikely that the next child will be a girl—and it was.

Simulation

The imitation of chance behavior, based on a model that accurately reflects the situation, is called a **simulation.** You already have some experience with simulations. In Chapter 1's "Hiring Discrimination—It Just Won't Fly!" Activity (page 5), you drew beads or slips of paper to imitate a random lottery to choose which pilots would become captains. Chapter 4's "Distracted Driving" Activity (page 249) asked you to shuffle and deal piles of cards to mimic the random assignment of subjects to treatments. The "1 in 6 wins" game that opened this chapter had you roll a die several times to simulate buying 20-ounce sodas and looking under the cap. These simulations involved different chance "devices"—beads, slips of paper, cards, or dice. But the same basic strategy was followed in all three simulations. We can summarize this strategy using our familiar four-step process: State, Plan, Do, Conclude.

PERFORMING A SIMULATION

STEP 4

State: Ask a question of interest about some chance process.

Plan: Describe how to use a chance device to imitate one repetition of the process. Tell what you will record at the end of each repetition.

Do: Perform many repetitions of the simulation.

Conclude: Use the results of your simulation to answer the question of interest.

The following table summarizes this four-step process for each of our previous simulations.

	Hiring Discrimination	Distracted Driving	"1 In 6 Wins" Game
State	How likely is it that a fair lottery would result in 5 or more female pilots being selected from an initial pool of 15 male and 10 female pilots?	Is it plausible that just by the luck of the random assignment of 48 subjects into two groups of 24 each, 12 of the 15 subjects who were going to miss the rest area anyway ended up in the cell-phone group, while only 3 of the 15 were assigned to the passenger conversation group?	What's the probability that 3 or more of 7 people who buy a 20-ounce bottle of soda win a prize if each bottle has a 1/6 chance of saying, "You're a winner!"?
Plan	Prepare a bag with 25 beads (15 of one color and 10 of another) or 25 identical slips of paper (15 labeled "M" and 10 labeled "F"). Mix well. Then without looking, remove 8 beads/slips from the bag. Record the number of female pilots selected.	• *Using index cards:* Write "Stop" on 33 cards and "Don't" on 15 cards. • *Using playing cards:* Remove jokers, specialty cards, the ace of spades, and three 2s from the deck. J, Q, K, A = miss rest area 2 through 10 = stop at rest area Shuffle well and deal two piles of 24 cards—the cell-phone group and the passenger group. Record the number of drivers who miss the rest area in the cell-phone group.	Use a six-sided die to determine the outcome for each person's bottle of soda. 6 = wins a prize 1 to 5 = no prize Roll the die seven times, once for each person. Record the number of people who win a prize.
Do	Have each student in class do this several times.	Have each pair of students repeat the process several times.	Have each student perform several repetitions.
Conclude	In 100 repetitions, 18 yielded 5 or more female pilots. So about 18% of the time, a fair lottery would choose at least 5 female pilots to become captains. It seems plausible that the company carried out a fair lottery.	In 300 repetitions of the random assignment, there were only 2 times when 12 or more drivers who missed the rest area ended up in the cell-phone group. That's less than 1% of the time! It doesn't seem plausible that the random assignment is the explanation for the difference between the groups.	Out of 125 total repetitions of the simulation, there were 15 times when three or more of the seven people won a prize. So our estimate of the probability is 15/125, or about 12%. It seems plausible that the company is telling the truth.

So far, we have used physical devices for our simulations. Using random numbers from Table D or technology is another option, as the following examples illustrate.

EXAMPLE Golden Ticket Parking Lottery

Simulations with Table D

At a local high school, 95 students have permission to park on campus. Each month, the student council holds a "golden ticket parking lottery" at a school assembly. The two lucky winners are given reserved parking spots next to the school's main entrance. Last month, the winning tickets were drawn by a student council member from the AP® Statistics class. When both golden tickets went to members of that same class, some people thought the lottery had been rigged. There are 28 students in the AP® Statistics class, all of whom are eligible to park on campus. Design and carry out a simulation to decide whether it's plausible that the lottery was carried out fairly.

STATE: *What's the probability that a fair lottery would result in two winners from the AP® Statistics class?*

PLAN: *We'll use Table D to simulate choosing the golden ticket lottery winners. Because there are 95 eligible students in the lottery, we'll label the students in the AP® Statistics class from 01 to 28,*

and the remaining students from 29 to 95. Numbers from 96 to 00 will be skipped. Moving left to right across the row, we'll look at pairs of digits until we come across two *different* labels from 01 to 95. The two students with these labels will win the prime parking spaces. We will record the number of winners from the AP® Statistics class.

DO: Let's perform many repetitions of our simulation. We'll use Table D starting at line 139. The digits from that row are shown below. We have drawn vertical bars to separate the pairs of digits. Underneath each pair, we have marked a √ if the chosen student is in the AP® Statistics class, × if the student is not in the class, and "skip" if the number isn't between 01 and 95 or if that student was already selected. (Note that if the consecutive "70" labels had been in the same repetition, we would have skipped the second one.) We also recorded the number of students in the AP® Statistics class for each repetition.

Rep 1	Rep 2	Rep 3	Rep 4	Rep 5	Rep 6	Rep 7	Rep 8	Rep 9	Rep 10
55∣58	89∣94	04∣70	70∣84	10∣98∣43	56∣35	69∣34	48∣39	45∣17	19∣12
× ×	× ×	✓ ×	× ×	✓ skip ×	× ×	× ×	× ×	× ✓	✓ ✓
0	0	1	0	1	0	0	0	1	2

In the first 10 repetitions, there was one time when the two winners were both from the AP® Statistics class. But 10 isn't many repetitions of the simulation. Continuing where we left off,

Rep 11	Rep 12	Rep 13	Rep 14	Rep 15	Rep 16	Rep 17	Rep 18	Rep 19	Rep 20
97∣51∣32	58∣13	04∣84	51∣44	72∣32	18∣19	40∣00∣36	00∣24∣28	96∣76∣73	59∣64
skip × ×	× ✓	✓ ×	× ×	× ×	✓ ✓	× skip ×	skip ✓ ✓	skip × ×	× ×
0	1	1	0	0	2	0	2	0	0

So after 20 repetitions, there have been 3 times when both winners were in the AP® Statistics class. If we keep going for 30 more repetitions (to bring our total to 50), we find 28 more "No" and 2 more "Yes" results. All totaled, that's 5 "Yes" and 45 "No" results.

CONCLUDE: In our simulation of a fair lottery, both winners came from the AP® Statistics class in 10% of the repetitions. So about 1 in every 10 times the student council holds the golden ticket lottery, this will happen just by chance. It seems plausible that the lottery was conducted fairly.

AP® EXAM TIP On the AP® exam, you may be asked to describe how you will perform a simulation using rows of random digits. If so, provide a clear enough description of your simulation process for the reader to get the same results you did from *only* your written explanation.

For Practice *Try Exercise* 19

In the previous example, we could have saved a little time by using `randInt(1,95)` repeatedly instead of Table D (so we wouldn't have to worry about numbers 96 to 00). We'll take this alternate approach in the next example.

EXAMPLE NASCAR Cards and Cereal Boxes

Simulations with technology

In an attempt to increase sales, a breakfast cereal company decides to offer a NASCAR promotion. Each box of cereal will contain a collectible card featuring one of these NASCAR drivers: Jeff Gordon, Dale Earnhardt, Jr., Tony Stewart, Danica Patrick, or Jimmie Johnson. The company says that each of the 5 cards is equally likely to appear in any box of cereal. A NASCAR fan decides to keep buying boxes of the cereal until she has all 5 drivers' cards. She is surprised when it takes her 23 boxes to get the full set of cards. Should she be surprised? Design and carry out a simulation to help answer this question.

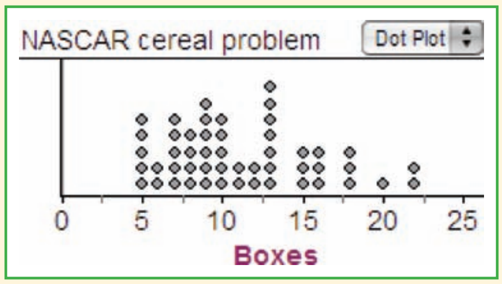
STATE: What is the probability that it will take 23 or more boxes to get a full set of 5 NASCAR collectible cards?

PLAN: We need five numbers to represent the five possible cards. Let's let 1 = Jeff Gordon; 2 = Dale Earnhardt, Jr.; 3 = Tony Stewart; 4 = Danica Patrick; and 5 = Jimmie Johnson. We'll use randInt(1,5) to simulate buying one box of cereal and looking at which card is inside. Because we want a full set of cards, we'll keep pressing Enter until we get all five of the labels from 1 to 5. We'll record the number of boxes that we had to open.

DO: It's time to perform many repetitions of the simulation. Here are our first few results:

Rep 1: 3 5 2 1 5 2 3 5 4 9 boxes
Rep 2: 5 1 2 5 1 4 1 4 1 2 2 2 4 4 5 3 16 boxes
Rep 3: 5 5 5 2 4 1 2 1 5 3 10 boxes
Rep 4: 4 3 5 3 5 1 1 1 5 3 1 5 4 5 2 15 boxes
Rep 5: 3 3 2 2 1 2 4 3 3 4 2 2 3 3 3 2 3 3 4 2 2 5 22 boxes

The Fathom dotplot shows the number of boxes we had to buy in 50 repetitions of the simulation.



CONCLUDE: We never had to buy more than 22 boxes to get the full set of NASCAR drivers' cards in 50 repetitions of our simulation. So our estimate of the probability that it takes 23 or more boxes to get a full set is roughly 0. The NASCAR fan should be surprised by how many boxes she had to buy.

For Practice *Try Exercise* **25**

In the golden ticket lottery example, we ignored repeated numbers from 01 to 95 within a given repetition. That's because the chance process involved sampling students *without* replacement. In the NASCAR example, we allowed repeated numbers from 1 to 5 in a given repetition. That's because we are selecting a small number of cards from a very large population of cards in thousands of cereal boxes. So the probability of getting, say, a Danica Patrick card in the next box of cereal is still very close to 1/5 even if we have already selected a Danica Patrick card.

THINK ABOUT IT

What *don't* these simulations tell us? For the golden ticket parking lottery, we concluded that it's plausible the drawing was done fairly. Does that mean the lottery *was* conducted fairly? Not necessarily. All we did was estimate that the probability of getting two winners from the AP® Statistics class was about 10% *if* the drawing was fair. So the result isn't unlikely enough to convince us that the lottery was rigged. What about the cereal box simulation? It took our NASCAR fan 23 boxes to complete the set of 5 cards. Does that mean the company didn't

tell the truth about how the cards were distributed? Not necessarily. Our simulation says that it's very unlikely for someone to have to buy 23 boxes to get a full set *if* each card is equally likely to appear in a box of cereal. The evidence suggests that the company's statement is incorrect. It's still possible, however, that the NASCAR fan was just very unlucky.

CHECK YOUR UNDERSTANDING

1. Refer to the golden ticket parking lottery example. At the following month's school assembly, the two lucky winners were once again members of the AP® Statistics class. This raised suspicions about how the lottery was being conducted. How would you modify the simulation in the example to estimate the probability of getting two winners from the AP® Statistics class in back-to-back months just by chance?

2. Refer to the NASCAR and breakfast cereal example. What if the cereal company decided to make it harder to get some drivers' cards than others? For instance, suppose the chance that each card appears in a box of the cereal is Jeff Gordon, 10%; Dale Earnhardt, Jr., 30%; Tony Stewart, 20%; Danica Patrick, 25%; and Jimmie Johnson, 15%. How would you modify the simulation in the example to estimate the chance that a fan would have to buy 23 or more boxes to get the full set?

Section 5.1 Summary

- A chance process has outcomes that we cannot predict but that nonetheless have a regular distribution in very many repetitions. The **law of large numbers** says that the proportion of times that a particular outcome occurs in many repetitions will approach a single number. This long-run relative frequency of a chance outcome is its **probability.** A probability is a number between 0 (never occurs) and 1 (always occurs).
- Probabilities describe only what happens in the long run. Short runs of random phenomena like tossing coins or shooting a basketball often don't look random to us because they do not show the regularity that emerges only in very many repetitions.
- A **simulation** is an imitation of chance behavior, most often carried out with random numbers. To perform a simulation, follow the four-step process:

STEP 4

STATE: Ask a question of interest about some chance process.

PLAN: Describe how to use a chance device to imitate one repetition of the process. Tell what you will record at the end of each repetition.

DO: Perform many repetitions of the simulation.

CONCLUDE: Use the results of your simulation to answer the question of interest.

Section 5.1 Exercises

1. **Liar, liar!** Sometimes police use a lie detector (also known as a polygraph) to help determine whether a suspect is telling the truth. A lie detector test isn't foolproof—sometimes it suggests that a person is lying when he or she is actually telling the truth (a "false positive"). Other times, the test says that the suspect is being truthful when the person is actually lying (a "false negative"). For one brand of polygraph machine, the probability of a false positive is 0.08.

(a) Interpret this probability as a long-run relative frequency.

(b) Which is a more serious error in this case: a false positive or a false negative? Justify your answer.

2. **Mammograms** Many women choose to have annual mammograms to screen for breast cancer after age 40. A mammogram isn't foolproof. Sometimes the test suggests that a woman has breast cancer when she really doesn't (a "false positive"). Other times the test says that a woman doesn't have breast cancer when she actually does (a "false negative"). Suppose the false negative rate for a mammogram is 0.10.

(a) Interpret this probability as a long-run relative frequency.

(b) Which is a more serious error in this case: a false positive or a false negative? Justify your answer.

3. **Genetics** Suppose a married man and woman both carry a gene for cystic fibrosis but don't have the disease themselves. According to the laws of genetics, the probability that their first child will develop cystic fibrosis is 0.25.

(a) Explain what this probability means.

(b) If the couple has 4 children, is one of them guaranteed to get cystic fibrosis? Explain.

4. **Texas hold 'em** In the popular Texas hold 'em variety of poker, players make their best five-card poker hand by combining the two cards they are dealt with three of five cards available to all players. You read in a book on poker that if you hold a pair (two cards of the same rank) in your hand, the probability of getting four of a kind is 88/1000.

(a) Explain what this probability means.

(b) If you play 1000 such hands, will you get four of a kind in exactly 88 of them? Explain.

5. **Spinning a quarter** With your forefinger, hold a new quarter (with a state featured on the reverse) upright, on its edge, on a hard surface. Then flick it with your other forefinger so that it spins for some time before it falls and comes to rest. Spin the coin a total of 25 times, and record the results.

(a) What's your estimate for the probability of heads? Why?

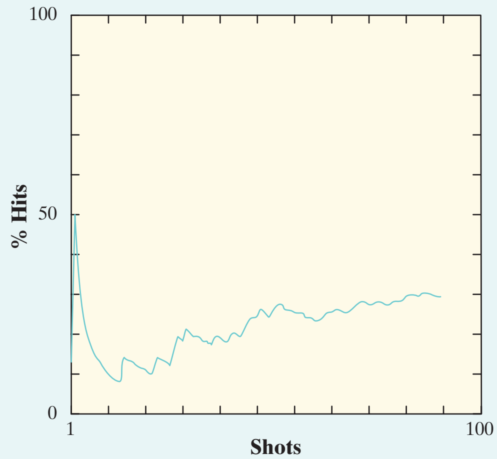
(b) Explain how you could get an even better estimate.

6. **Nickels falling over** You may feel it's obvious that the probability of a head in tossing a coin is about 1/2 because the coin has two faces. Such opinions are not always correct. Stand a nickel on edge on a hard, flat surface. Pound the surface with your hand so that the nickel falls over. Do this 25 times, and record the results.

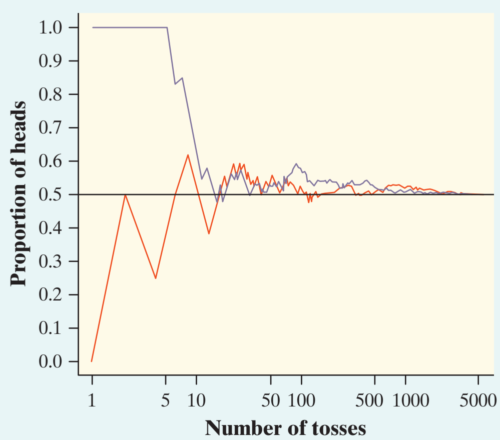
(a) What's your estimate for the probability that the coin falls heads up? Why?

(b) Explain how you could get an even better estimate.

7. **Free throws** The figure below shows the results of a virtual basketball player shooting several free throws. Explain what this graph says about chance behavior in the short run and long run.



8. **Keep on tossing** The figure below shows the results of two different sets of 5000 coin tosses. Explain what this graph says about chance behavior in the short run and the long run.



9. **Due for a hit** A very good professional baseball player gets a hit about 35% of the time over an entire season. After the player failed to hit safely in six straight at-bats, a TV commentator said, "He is due for a hit by the law of averages." Is that right? Why?

10. **Cold weather coming** A TV weather man, predicting a colder-than-normal winter, said, "First, in looking at the past few winters, there has been a lack of really cold weather. Even though we are not supposed to use the law of averages, we are due." Do you think that "due by the law of averages" makes sense in talking about the weather? Why or why not?

11. **Playing "Pick 4"** The Pick 4 games in many state lotteries announce a four-digit winning number each day. You can think of the winning number as a four-digit group from a table of random digits. You win (or share) the jackpot if your choice matches the winning number. The winnings are divided among all players who matched the winning number. That suggests a way to get an edge.

(a) The winning number might be, for example, either 2873 or 9999. Explain why these two outcomes have exactly the same probability.

(b) If you asked many people whether 2873 or 9999 is more likely to be the randomly chosen winning number, most would favor one of them. Use the information in this section to say which one and to explain why. How might this affect the four-digit number you would choose?

12. **An unenlightened gambler**

(a) A gambler knows that red and black are equally likely to occur on each spin of a roulette wheel. He observes five consecutive reds occur and bets heavily on black at the next spin. Asked why, he explains that black is "due by the law of averages." Explain to the gambler what is wrong with this reasoning.

(b) After hearing you explain why red and black are still equally likely after five reds on the roulette wheel, the gambler moves to a poker game. He is dealt five straight red cards. He remembers what you said and assumes that the next card dealt in the same hand is equally likely to be red or black. Is the gambler right or wrong, and why?

13. **Free throws** A basketball player has probability 0.75 of making a free throw. Explain how you would use each chance device to simulate one free throw by the player.

(a) A standard deck of playing cards

(b) Table D of random digits

(c) A calculator or computer's random integer generator

14. **Stoplight** On her drive to work every day, Ilana passes through an intersection with a traffic light. The light has probability 1/3 of being green when she gets to the intersection. Explain how you would use each chance device to simulate whether the light is red or green on a given day.

(a) A six-sided die

(b) Table D of random digits

(c) A calculator or computer's random integer generator

15. **Simulation blunders** Explain what's wrong with each of the following simulation designs.

(a) A roulette wheel has 38 colored slots—18 red, 18 black, and 2 green. To simulate one spin of the wheel, let numbers 00 to 18 represent red, 19 to 37 represent black, and 38 to 40 represent green.

(b) About 10% of U.S. adults are left-handed. To simulate randomly selecting one adult at a time until you find a left-hander, use two digits. Let 00 to 09 represent being left-handed and 10 to 99 represent being right-handed. Move across a row in Table D, two digits at a time, skipping any numbers that have already appeared, until you find a number between 00 and 09. Record the number of people selected.

16. **Simulation blunders** Explain what's wrong with each of the following simulation designs.

(a) According to the Centers for Disease Control and Prevention, about 36% of U.S. adults were obese in 2012. To simulate choosing 8 adults at random and seeing how many are obese, we could use two digits. Let 00 to 35 represent obese and 36 to 99 represent not obese. Move across a row in Table D, two digits at a time, until you find 8 distinct numbers (no repeats). Record the number of obese people selected.

(b) Assume that the probability of a newborn being a boy is 0.5. To simulate choosing a random sample of 9 babies who were born at a local hospital today and observing their gender, use one digit. Use `randInt(0,9)` on your calculator to determine how many babies in the sample are male.

17. **Is this valid?** Determine whether each of the following simulation designs is valid. Justify your answer.

(a) According to a recent poll, 75% of American adults regularly recycle. To simulate choosing a random sample of 100 U.S. adults and seeing how many of them recycle, roll a 4-sided die 100 times. A result of 1, 2, or 3 means the person recycles; a 4 means that the person doesn't recycle.

(b) An archer hits the center of the target with 60% of her shots. To simulate having her shoot 10 times, use a coin. Flip the coin once for each of the 10 shots. If it lands heads, then she hits the center of the target. If the coin lands tails, she doesn't.

18. **Is this valid?** Determine whether each of the following simulation designs is valid. Justify your answer.

(a) According to a recent survey, 50% of people aged 13 and older in the United States are addicted to texting. To simulate choosing a random sample of 20 people in this population and seeing how many of them are addicted to texting, use a deck of cards. Shuffle the deck well, and then draw one card at a time. A red card means that person is addicted to texting; a black card means he isn't. Continue until you have drawn 20 cards (without replacement) for the sample.

(b) A tennis player gets 95% of his serves in play during practice (that is, the ball doesn't go out of bounds). To simulate the player hitting 5 serves, look at 5 pairs of digits going across a row in Table D. If the number is between 00 and 94, the serve is in; numbers between 95 and 99 indicate that the serve is out.

19. **Airport security** The Transportation Security Administration (TSA) is responsible for airport safety. On some flights, TSA officers randomly select passengers for an extra security check prior to boarding. One such flight had 76 passengers—12 in first class and 64 in coach class. Some passengers were surprised when none of the 10 passengers chosen for screening were seated in first class. We can use a simulation to see if this result is likely to happen by chance.

pg 296

(a) State the question of interest using the language of probability.

(b) How would you use random digits to imitate one repetition of the process? What variable would you measure?

(c) Use the line of random digits below to perform one repetition. Copy these digits onto your paper. Mark directly on or above them to show how you determined the outcomes of the chance process.

71487 09984 29077 14863 61683 47052 62224 51025

(d) In 100 repetitions of the simulation, there were 15 times when none of the 10 passengers chosen was seated in first class. What conclusion would you draw?

20. **Scrabble** In the game of Scrabble, each player begins by drawing 7 tiles from a bag containing 100 tiles. There are 42 vowels, 56 consonants, and 2 blank tiles in the bag. Cait chooses her 7 tiles and is surprised to discover that all of them are vowels. We can use a simulation to see if this result is likely to happen by chance.

(a) State the question of interest using the language of probability.

(b) How would you use random digits to imitate one repetition of the process? What variable would you measure?

(c) Use the line of random digits below to perform one repetition. Copy these digits onto your paper. Mark directly on or above them to show how you determined the outcomes of the chance process.

00694 05977 19664 65441 20903 62371 22725 53340

(d) In 1000 repetitions of the simulation, there were 2 times when all 7 tiles were vowels. What conclusion would you draw?

21. **The birthday problem** What's the probability that in a randomly selected group of 30 unrelated people, at least two have the same birthday? Let's make two assumptions to simplify the problem. First, we'll ignore the possibility of a February 29 birthday. Second, we assume that a randomly chosen person is equally likely to be born on each of the remaining 365 days of the year.

(a) How would you use random digits to imitate one repetition of the process? What variable would you measure?

(b) Use technology to perform 5 repetitions. Record the outcome of each repetition.

(c) Would you be surprised to learn that the theoretical probability is 0.71? Why or why not?

22. **Monty Hall problem** In *Parade* magazine, a reader posed the following question to Marilyn vos Savant and the "Ask Marilyn" column:

> Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say #1, and the host, who knows what's behind the doors, opens another door, say #3, which has a goat. He says to you, "Do you want to pick door #2?" Is it to your advantage to switch your choice of doors?[4]

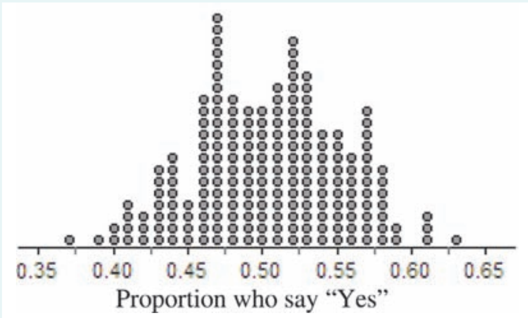
The game show in question was *Let's Make a Deal* and the host was Monty Hall. Here's the first part of Marilyn's response: "Yes; you should switch. The first door has a 1/3 chance of winning, but the second door has a 2/3 chance." Thousands of readers wrote to Marilyn to disagree with her answer. But she held her ground.

(a) Use an online *Let's Make a Deal* applet to perform at least 50 repetitions of the simulation. Record whether you stay or switch (try to do each about half the time) and the outcome of each repetition.

(b) Do you agree with Marilyn or her readers? Explain.

23. **Recycling** Do most teens recycle? To find out, an AP® Statistics class asked an SRS of 100 students at their school whether they regularly recycle. In the sample, 55 students said that they recycle. Is this convincing evidence that more than half of the students at the school would say they regularly recycle? The Fathom dotplot below shows the results of

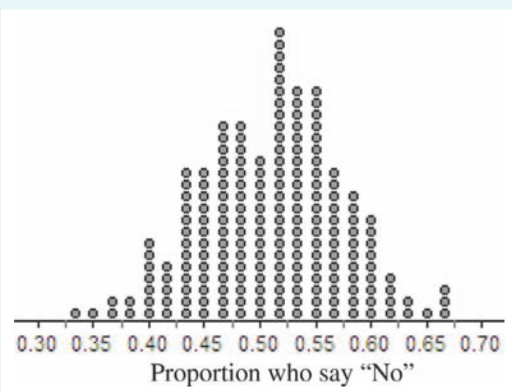
taking 200 SRSs of 100 students from a population in which the true proportion who recycle is 0.50.



(a) Explain why the sample result does not give convincing evidence that more than half of the school's students recycle.

(b) Suppose instead that 63 students in the class's sample had said "Yes." Explain why this result would give strong evidence that a majority of the school's students recycle.

24. **Brushing teeth, wasting water?** A recent study reported that fewer than half of young adults turn off the water while brushing their teeth. Is the same true for teenagers? To find out, a group of statistics students asked an SRS of 60 students at their school if they usually brush with the water off. In the sample, 27 students said "No." The Fathom dotplot below shows the results of taking 200 SRSs of 60 students from a population in which the true proportion who brush with the water off is 0.50.



(a) Explain why the sample result does not give convincing evidence that fewer than half of the school's students brush their teeth with the water off.

(b) Suppose instead that 18 students in the class's sample had said "No." Explain why this result would give strong evidence that fewer than 50% of the school's students brush their teeth with the water off.

25. **Color-blind men** About 7% of men in the United States have some form of red-green color blindness. Suppose we randomly select 4 U.S. adult males. What's the probability that at least one of them is red-green color-blind? Design and carry out a simulation to answer this question. Follow the four-step process.
pg 297
STEP 4

26. **Lotto** In the United Kingdom's Lotto game, a player picks six numbers from 1 to 49 for each ticket. Rosemary bought one ticket for herself. She had the lottery computer randomly select the six numbers. When the six winning numbers were drawn, Rosemary was surprised to find that none of these numbers appeared on the Lotto ticket she had bought. Should she be? Design and carry out a simulation to answer this question. Follow the four-step process.
STEP 4

27. **Color-blind men** Refer to Exercise 25. Suppose we randomly select one U.S. adult male at a time until we find one who is red-green color-blind. Should we be surprised if it takes us 20 or more men? Design and carry out a simulation to answer this question. Follow the four-step process.
STEP 4

28. **Scrabble** Refer to Exercise 20. About 3% of the time, the first player in Scrabble can "bingo" by playing all 7 tiles on the first turn. Should we be surprised if it takes 30 or more games for this to happen? Design and carry out a simulation to answer this question. Follow the four-step process.
STEP 4

29. **Random assignment** Researchers recruited 20 volunteers—8 men and 12 women—to take part in an experiment. They randomly assigned the subjects into two groups of 10 people each. To their surprise, 6 of the 8 men were randomly assigned to the same treatment. Should they be surprised? Design and carry out a simulation to estimate the probability that the random assignment puts 6 or more men in the same group. Follow the four-step process.
STEP 4

30. **Taking the train** According to New Jersey Transit, the 8:00 A.M. weekday train from Princeton to New York City has a 90% chance of arriving on time. To test this claim, an auditor chooses 6 weekdays at random during a month to ride this train. The train arrives late on 2 of those days. Does the auditor have convincing evidence that the company's claim isn't true? Design and carry out a simulation to estimate the probability that a train with a 90% chance of arriving on time each day would be late on 2 or more of 6 days. Follow the four-step process.
STEP 4

Multiple choice: Select the best answer for Exercises 31 to 36.

31. You read in a book about bridge that the probability that each of the four players is dealt exactly one ace is about 0.11. This means that

(a) in every 100 bridge deals, each player has one ace exactly 11 times.

(b) in 1 million bridge deals, the number of deals on which each player has one ace will be exactly 110,000.

(c) in a very large number of bridge deals, the percent of deals on which each player has one ace will be very close to 11%.

(d) in a very large number of bridge deals, the average number of aces in a hand will be very close to 0.11.

(e) If each player gets an ace in only 2 of the first 50 deals, then each player should get an ace in more than 11% of the next 50 deals.

32. If I toss a fair coin five times and the outcomes are TTTTT, then the probability that tails appears on the next toss is

(a) 0.5. (c) greater than 0.5. (e) 1.

(b) less than 0.5. (d) 0.

Exercises 33 to 35 refer to the following setting. A basketball player claims to make 47% of her shots from the field. We want to simulate the player taking sets of 10 shots, assuming that her claim is true.

33. To simulate the number of makes in 10 shot attempts, you would perform the simulation as follows:

(a) Use 10 random one-digit numbers, where 0–4 are a make and 5–9 are a miss.

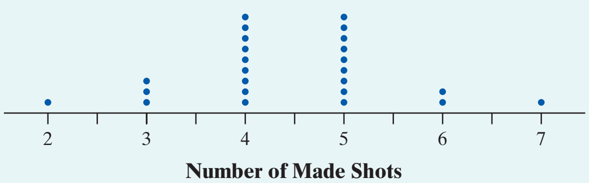
(b) Use 10 random two-digit numbers, where 00–46 are a make and 47–99 are a miss.

(c) Use 10 random two-digit numbers, where 00–47 are a make and 48–99 are a miss.

(d) Use 47 random one-digit numbers, where 0 is a make and 1–9 are a miss.

(e) Use 47 random two-digit numbers, where 00–46 are a make and 47–99 are a miss.

34. Twenty-five repetitions of the simulation were performed. The simulated number of makes in each set of 10 shots was recorded on the dotplot below. What is the approximate probability that a 47% shooter makes 5 or more shots in 10 attempts?



(a) 5/10 (b) 3/10 (c) 12/25 (d) 3/25 (e) 47/100

35. Suppose this player attempts 10 shots in a game and only makes 3 of them. Does this provide convincing evidence that she is less than a 47% shooter?

(a) Yes, because 3/10 (30%) is less than 47%.

(b) Yes, because she never made 47% of her shots in the simulation.

(c) No, because it is plausible that she would make 3 or fewer shots by chance alone.

(d) No, because the simulation was only repeated 25 times.

(e) No, because the distribution is approximately symmetric.

36. Ten percent of U.S. households contain 5 or more people. You want to simulate choosing a household at random and recording "Yes" if it contains 5 or more people. Which of these are correct assignments of digits for this simulation?

(a) Odd = Yes; Even = No

(b) 0 = Yes; 1–9 = No

(c) 0–5 = Yes; 6–9 = No

(d) 0–4 = Yes; 5–9 = No

(e) None of these

37. Are you feeling stressed? (4.1) A Gallup Poll asked whether people experienced stress "a lot of the day yesterday." About 41 percent said they did. Gallup's report said, "Results are based on telephone interviews conducted … Jan. 1–Dec. 31, 2012, with a random sample of 353,564 adults aged 18 and older."[5]

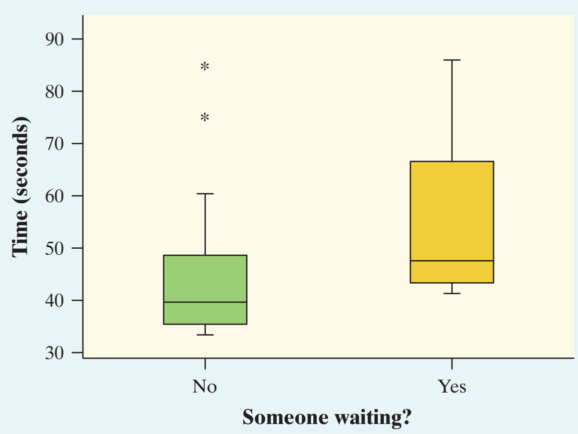
(a) Identify the population and the sample.

(b) Explain how undercoverage could lead to bias in this survey.

38. Waiting to park (1.3, 4.2) Do drivers take longer to leave their parking spaces when someone is waiting? Researchers hung out in a parking lot and collected some data. The graphs and numerical summaries below display information about how long it took drivers to exit their spaces.

(a) Write a few sentences comparing these distributions.

(b) Can we conclude that having someone waiting causes drivers to leave their spaces more slowly? Why or why not?



Descriptive Statistics: Time

Waiting	n	Mean	StDev	Minimum	Q_1	Median	Q_3	Maximum
No	20	44.42	14.10	33.76	35.61	39.56	48.48	84.92
Yes	20	54.11	14.39	41.61	43.41	47.14	66.44	85.97

5.2 Probability Rules

WHAT YOU WILL LEARN By the end of the section, you will be able to:

- Describe a probability model for a chance process.
- Use basic probability rules, including the complement rule and the addition rule for mutually exclusive events.
- Use a two-way table or Venn diagram to model a chance process and calculate probabilities involving two events.
- Use the general addition rule to calculate probabilities.

The idea of probability rests on the fact that chance behavior is predictable in the long run. In Section 5.1, we used simulation to imitate chance behavior. Do we always need to repeat a chance process many times to determine the probability of a particular outcome? Fortunately, the answer is no.

Probability Models

In Chapter 2, we saw that a Normal density curve could be used to model some distributions of data. In Chapter 3, we modeled linear relationships between two quantitative variables with a least-squares regression line. Now we're ready to develop a model for chance behavior.

Let's start with a very simple chance process: tossing a coin once. When we toss a coin, we can't know the outcome in advance. What *do* we know? We are willing to say that the outcome will be either heads or tails. We believe that each of these outcomes has probability 1/2. This description of coin tossing has two parts:

- A list of possible outcomes (the **sample space** S)
- A probability for each outcome

Such a description is the basis for a **probability model**. Here is the basic vocabulary we use.

DEFINITION: Sample space, probability model

The **sample space** S of a chance process is the set of all possible outcomes.

A **probability model** is a description of some chance process that consists of two parts: a sample space S and a probability for each outcome.

A sample space S can be very simple or very complex. When we toss a coin once, there are only two possible outcomes, heads and tails. We can write the sample space using set notation as $S = \{H, T\}$. When Gallup draws a random sample of 1523 adults and asks a survey question, the sample space contains all possible sets of responses from 1523 of the 235 million adults in the country. This S is extremely large. Each member of S lists the answers from one possible sample, which explains the term *sample space.*

Let's look at how to set up a probability model in a familiar setting—rolling a pair of dice.

EXAMPLE Roll the Dice

Building a probability model

Many board games involve rolling dice. Imagine rolling two fair, six-sided dice—one that's red and one that's green.

PROBLEM: *Give a probability model for this chance process.*

SOLUTION: *There are 36 possible outcomes when we roll two dice and record the number of spots showing on the up-faces. Figure 5.2 displays these outcomes. They make up the sample space S. If the dice are perfectly balanced, all 36 outcomes will be equally likely. That is, each of the 36 outcomes will come up on one-thirty-sixth of all rolls in the long run. So each outcome has probability 1/36.*

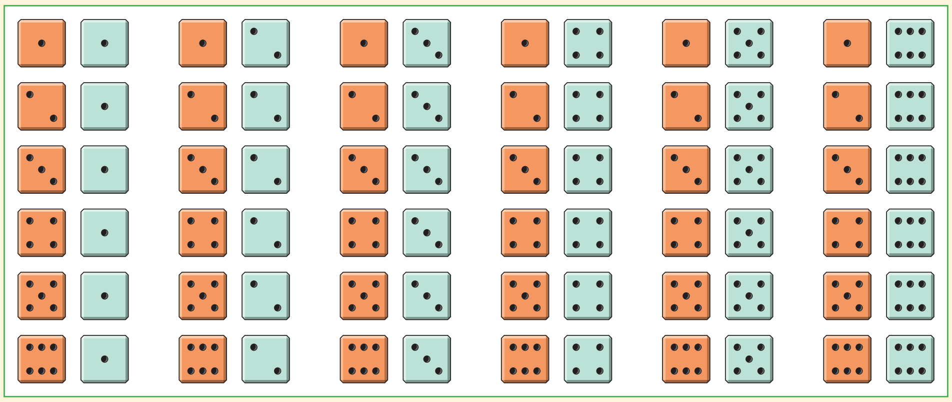


FIGURE 5.2 The 36 possible outcomes in rolling two dice. If the dice are carefully made, all of these outcomes have the same probability.

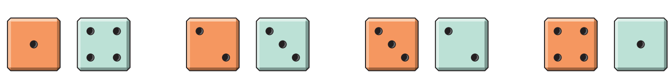
For Practice *Try Exercise* 39

A probability model does more than just assign a probability to each outcome. It allows us to find the probability of any collection of outcomes, which we call an **event.**

DEFINITION: Event

An **event** is any collection of outcomes from some chance process. That is, an event is a subset of the sample space. Events are usually designated by capital letters, like *A*, *B*, *C*, and so on.

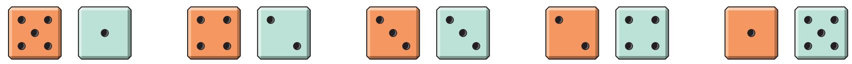
If A is any event, we write its probability as $P(A)$. In the dice-rolling example, suppose we define event A as "sum is 5." What's $P(A)$, the probability that event A occurs? There are four outcomes in event A:



Because each of these outcomes has probability 1/36, $P(A) = 4/36$. Now consider event B: sum is not 5. To find $P(B)$, we could list all the outcomes that make up

event B, but that would take a while. Fortunately, there's an easier way. Of the 36 equally likely outcomes in Figure 5.2, event A (sum is 5) occurs in 4 of them. So event A does *not* occur in 32 of these outcomes. Then $P(B) = P(\text{sum isn't } 5) = P(\text{not A}) = 32/36$. Notice that $P(\text{A}) + P(B) = 1$.

Let's consider one more event, which we'll call C: sum is 6. The outcomes in event C are



So $P(C) = 5/36$. What's the probability that we get a sum of 5 *or* 6, that is, $P(\text{A or C})$? Because these two events have no outcomes in common, we can add the probabilities of the individual events:

$$P(\text{sum is 5 or sum is 6}) = P(\text{sum is 5}) + P(\text{sum is 6}) = 4/36 + 5/36 = 9/36$$

In other words, $P(\text{A or C}) = P(\text{A}) + P(\text{C})$.

Basic Rules of Probability

Our dice-rolling scenario revealed some basic rules that any probability model must obey:

- *The probability of any event is a number between 0 and 1.* The probability of an event is the long-run proportion of repetitions on which that event occurs. Any proportion is a number between 0 and 1, so any probability is also a number between 0 and 1. An event with probability 0 never occurs, and an event with probability 1 occurs on every trial. An event with probability 0.5 occurs in half the trials in the long run.
- *All possible outcomes together must have probabilities that add up to 1.* Because some outcome must occur on every trial, the sum of the probabilities for all possible outcomes must be exactly 1.
- *If all outcomes in the sample space are equally likely, the probability that event A occurs can be found using the formula*

$$P(\text{A}) = \frac{\text{number of outcomes corresponding to event A}}{\text{total number of outcomes in sample space}}$$

- *The probability that an event does not occur is 1 minus the probability that the event does occur.* If an event occurs in (say) 70% of all trials, it fails to occur in the other 30%. The probability that an event occurs and the probability that it does not occur always add to 100%, or 1. (This explains why $P(\text{sum isn't } 5) = 1 - P(\text{sum is } 5)$ in the dice-rolling example.) We refer to the event "not A" as the **complement** of A and denote it by A^C.
- *If two events have no outcomes in common, the probability that one or the other occurs is the sum of their individual probabilities.* If one event occurs in 40% of all trials, a different event occurs in 25% of all trials, and the two can never occur together, then one or the other occurs on 65% of all trials because 40% + 25% = 65%. When two events have no outcomes in common, we refer to them as **mutually exclusive** or **disjoint**.

DEFINITION: Mutually exclusive (disjoint)

Two events A and B are **mutually exclusive (disjoint)** if they have no outcomes in common and so can never occur together—that is, if $P(A \text{ and } B) = 0$.

We can summarize the basic probability rules more concisely in symbolic form.

BASIC PROBABILITY RULES

- For any event A, $0 \le P(A) \le 1$.
- If S is the sample space in a probability model, $P(S) = 1$.
- In the case of *equally likely* outcomes,

$$P(A) = \frac{\text{number of outcomes corresponding to event A}}{\text{total number of outcomes in sample space}}$$

- **Complement rule:** $P(A^C) = 1 - P(A)$.
- **Addition rule for mutually exclusive events:** If A and B are mutually exclusive, $P(A \text{ or } B) = P(A) + P(B)$.

The earlier dice-rolling example involved equally likely outcomes. Here's an example that illustrates use of the basic probability rules when the outcomes of a chance process are not equally likely.

EXAMPLE Distance Learning

Applying probability rules

Distance-learning courses are rapidly gaining popularity among college students. Randomly select an undergraduate student who is taking a distance-learning course for credit, and record the student's age. Here is the probability model:[6]

Age group (yr):	18 to 23	24 to 29	30 to 39	40 or over
Probability:	0.57	0.17	0.14	0.12

PROBLEM:

(a) Show that this is a legitimate probability model.

(b) Find the probability that the chosen student is not in the traditional college age group (18 to 23 years).

SOLUTION:

(a) The probability of each outcome is a number between 0 and 1, and the probabilities of all the possible outcomes add to 1, so this is a legitimate probability model.

(b) There are two ways to find this probability. By the complement rule,

$$P(\text{not 18 to 23 years}) = 1 - P(\text{18 to 23 years}) = 1 - 0.57 = 0.43$$

That is, if 57% of distance learners are 18 to 23 years old, then the remaining 43% are not in this age group.

Using the addition rule for mutually exclusive events,

$$\begin{aligned} P(\text{not 18 to 23 years}) &= P(\text{24 to 29 years}) + P(\text{30 to 39 years}) + P(\text{40 years or over}) \\ &= 0.17 + 0.14 + 0.12 = 0.43 \end{aligned}$$

There is a 43% chance that the chosen student is not in the traditional college age group.

For Practice *Try Exercise* **45**

CHECK YOUR UNDERSTANDING

Choose an American adult at random. Define two events:

A = the person has a cholesterol level of 240 milligrams per deciliter of blood (mg/dl) or above (high cholesterol)
B = the person has a cholesterol level of 200 to 239 mg/dl (borderline high cholesterol)
According to the American Heart Association, $P(A) = 0.16$ and $P(B) = 0.29$.

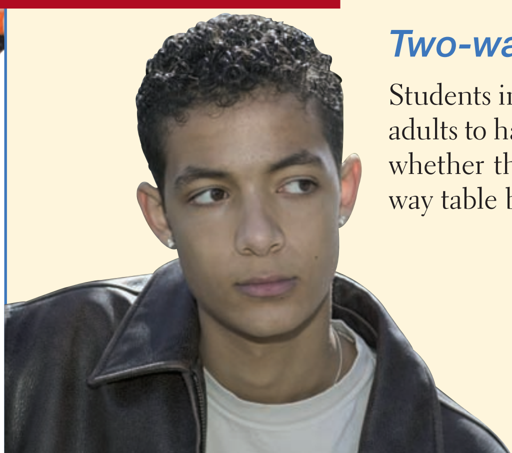
1. Explain why events A and B are mutually exclusive.
2. Say in plain language what the event "A or B" is. What is $P(A \text{ or } B)$?
3. If C is the event that the person chosen has normal cholesterol (below 200 mg/dl), what's $P(C)$?

Two-Way Tables, Probability, and the General Addition Rule

When we're trying to find probabilities involving two events, a two-way table can display the sample space in a way that makes probability calculations easier.

EXAMPLE Who Has Pierced Ears?

Two-way tables and probability



Students in a college statistics class wanted to find out how common it is for young adults to have their ears pierced. They recorded data on two variables—gender and whether the student had a pierced ear—for all 178 people in the class. The two-way table below displays the data.

	Gender		
Pierced Ears?	Male	Female	**Total**
Yes	19	84	**103**
No	71	4	**75**
Total	**90**	**88**	**178**

PROBLEM: Suppose we choose a student from the class at random. Find the probability that the student

(a) has pierced ears.

(b) is male and has pierced ears.

(c) is male or has pierced ears.

SOLUTION: We'll define events A: is male and B: has pierced ears.

(a) Because each of the 178 students in the class is equally likely to be chosen, and there are 103 students with pierced ears, $P(\text{pierced ears}) = P(B) = 103/178$.

(b) We want to find $P(\text{male } \textit{and} \text{ pierced ears})$, that is, $P(A \textit{ and } B)$. Looking at the intersection of the "Male" column and "Yes" row, we see that there are 19 males with pierced ears. So $P(\text{male and pierced ears}) = P(A \text{ and } B) = 19/178$.

(c) This time, we're interested in $P(\text{male } \textit{or} \text{ pierced ears})$, that is, $P(A \textit{ or } B)$. (Note the mathematical use of the word *or* here—the person could be a male or have pierced ears or both.) From the two-way table,

we see that there are 90 males in the class, so $P(A) = 90/178$. Can we just add $P(A)$ to $P(B)$ to get the correct answer? No! These two events are *not* mutually exclusive, because there are 19 males with pierced ears. (If we did add the two probabilities, we'd get $90/178 + 103/178 = 193/178$, which is clearly wrong, because the probability is bigger than 1!) From the two-way table, we see that there are $19 + 71 + 84 = 174$ students who are male *or* have pierced ears. So $P(A \text{ or } B) = 174/178$.

For Practice *Try Exercise* **49**

When we found the probability of getting a male with pierced ears in the example, we could have described this as either $P(A \text{ and } B)$ or $P(B \text{ and } A)$. Why? Because "A and B" describes the same event as "B and A." Likewise, $P(A \text{ or } B)$ is the same as $P(B \text{ or } A)$. *Don't get so caught up in the notation that you lose sight of what's really happening!*

The previous example revealed two important facts about finding the probability $P(\text{A or B})$ when the two events are not mutually exclusive. First, the use of the word "or" in probability questions is different from that in everyday life. If someone says, "I'll either watch a movie or go to the football game," that usually means they'll do one thing or the other, but not both. In statistics, "A or B" could mean one or the other or both. Second, we can't use the addition rule for mutually exclusive events unless two events have no outcomes in common.

If events A and B are *not* mutually exclusive, they can occur together. The probability that one or the other occurs is then *less* than the sum of their probabilities. As Figure 5.3 illustrates, outcomes common to both are counted twice when we add probabilities.

Outcomes here are double-counted by $P(A) + P(B)$

$P(A \text{ and } B) = \dfrac{19}{178}$

	Gender			
Pierced Ears?	**Male**	**Female**	**Total**	
Yes	19	84	**103**	$P(B) = \dfrac{103}{178}$
No	71	4	**75**	
Total	**90**	**88**	**178**	
	$P(A) = \dfrac{90}{178}$			

FIGURE 5.3 Two-way table showing events A and B from the pierced-ears example. These events are *not* mutually exclusive, so we can't find $P(A \text{ or } B)$ by just adding the probabilities of the two events.

We can fix the double-counting problem illustrated in the two-way table by subtracting the probability $P(\text{A and B})$ from the sum. That is,

$$P(\text{A or B}) = P(\text{A}) + P(\text{B}) - P(\text{A and B})$$

This result is known as the **general addition rule.** Let's check that it works for the pierced-ears example:

$$\begin{aligned} P(\text{A or B}) &= P(\text{A}) + P(\text{B}) - P(\text{A and B}) \\ &= 90/178 + 103/178 - 19/178 \\ &= 174/178 \end{aligned}$$

This matches our earlier result.

GENERAL ADDITION RULE FOR TWO EVENTS

If A and B are any two events resulting from some chance process, then

$$P(\text{A or B}) = P(\text{A}) + P(\text{B}) - P(\text{A and B})$$

What happens if we use the general addition rule for two mutually exclusive events A and B? In that case, $P(A \text{ and } B) = 0$, and the formula reduces to $P(A \text{ or } B) = P(A) + P(B)$. In other words, the addition rule for mutually exclusive events is just a special case of the general addition rule.

You might be wondering whether there is also a rule for finding $P(A \text{ and } B)$. There is, but it's not quite as intuitive. Stay tuned for that later.

CHECK YOUR UNDERSTANDING

Sometimes it's easier to designate events with letters that relate to the context, like F for "face card" and H for "heart."

A standard deck of playing cards (with jokers removed) consists of 52 cards in four suits—clubs, diamonds, hearts, and spades. Each suit has 13 cards, with denominations ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, jack, queen, and king. The jack, queen, and king are referred to as "face cards." Imagine that we shuffle the deck thoroughly and deal one card. Let's define events F: getting a face card and H: getting a heart.

1. Make a two-way table that displays the sample space.
2. Find $P(F \text{ and } H)$.
3. Explain why $P(F \text{ or } H) \neq P(F) + P(H)$. Then use the general addition rule to find $P(F \text{ or } H)$.

Venn Diagrams and Probability

We have already seen that two-way tables can be used to illustrate the sample space of a chance process involving two events. So can **Venn diagrams**. Because Venn diagrams have uses in other branches of mathematics, some standard vocabulary and notation have been developed.

- We introduced the *complement* of an event earlier. In Figure 5.4(a), the complement A^C contains exactly the outcomes that are not in A.
- The events A and B in Figure 5.4(b) are *mutually exclusive* (*disjoint*) because they do not overlap; that is, they have no outcomes in common.

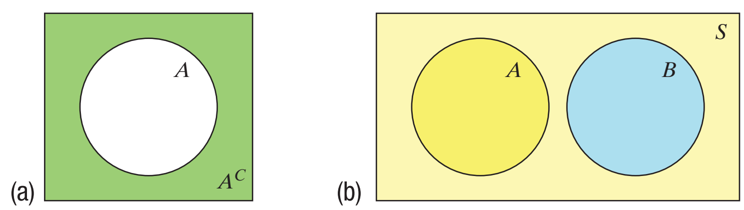


FIGURE 5.4 Venn diagrams showing: (a) event A and its complement A^C and (b) mutually exclusive (disjoint) events A and B.

- Figure 5.5(a) shows the event "A *and* B." You can see why this event is also called the **intersection** of A and B. The corresponding notation is $A \cap B$.
- The event "A or B" is shown in Figure 5.5(b). This event is also known as the **union** of A and B. The corresponding notation is $A \cup B$.

Here's a way to keep the symbols straight: $\cup$ for **u**nion; $\cap$ for **in**tersection.

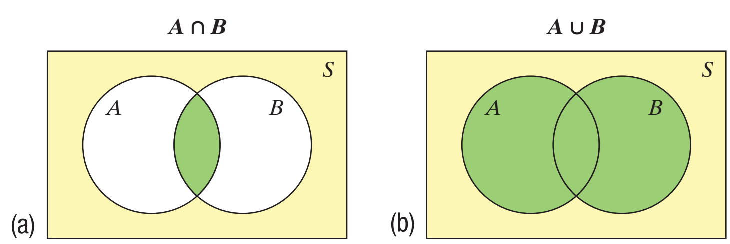


FIGURE 5.5 Venn diagrams showing (a) the intersection and (b) the union of events A and B.

EXAMPLE Who Has Pierced Ears?

Understanding Venn diagrams

In the preceding example, we looked at data from a survey on gender and ear piercings for a large group of college students. The chance process came from selecting a student in the class at random. Our events of interest were A: is male and B: has pierced ears. Here is the two-way table that summarizes the sample space:

	Gender	
Pierced Ears?	**Male**	**Female**
Yes	19	84
No	71	4

How would we construct a Venn diagram that displays the information in the two-way table?

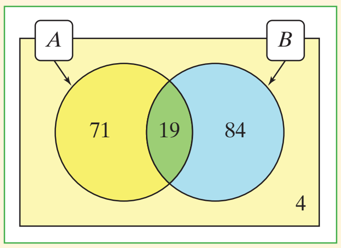


FIGURE 5.6 Completed Venn diagram for the large group of college students. The circles represent the two events A: is male and B: has pierced ears.

There are four distinct regions in the Venn diagram shown in Figure 5.6. These regions correspond to the four cells in the two-way table. We can describe this correspondence in tabular form as follows:

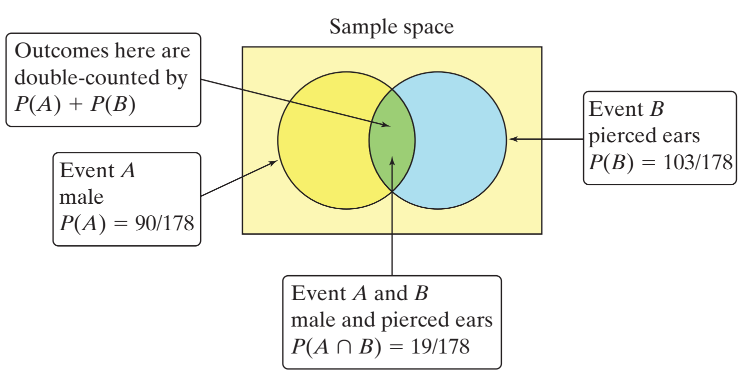
Region in Venn diagram	In words	In symbols	Count
In the intersection of two circles	Male and pierced ears	$A \cap B$	19
Inside circle A, outside circle B	Male and no pierced ears	$A \cap B^C$	71
Inside circle B, outside circle A	Female and pierced ears	$A^C \cap B$	84
Outside both circles	Female and no pierced ears	$A^C \cap B^C$	4

We have added the appropriate counts of students to the four regions in Figure 5.6.

With this new notation, we can rewrite the general addition rule in symbols as

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

This Venn diagram shows why the formula works in the pierced-ears example.



The following example ties all this together.

EXAMPLE Who Reads the Paper?

Venn diagrams, two-way tables, and probability

In an apartment complex, 40% of residents read *USA Today*. Only 25% read the *New York Times*. Five percent of residents read both papers. Suppose we select a resident of the apartment complex at random and record which of the two papers the person reads.

PROBLEM:

(a) Make a two-way table that displays the sample space of this chance process.

(b) Construct a Venn diagram to represent the outcomes of this chance process.

(c) Find the probability that the person reads at least one of the two papers.

(d) Find the probability that the person doesn't read either paper.

SOLUTION: We'll define events A: reads *USA Today* and B: reads *New York Times*. From the problem statement, we know that $P(A) = 0.40$, $P(B) = 0.25$, and $P(A \cap B) = 0.05$.

(a) We can enter the value 0.40 as the total for the "Yes" column, 0.25 as the total for the "Yes" row, 0.05 in the "Yes, Yes" cell, and 1 as the grand total in the two-way table shown here. This gives us enough information to fill in the empty cells of the table, starting with the missing row total for "No" ($1 - 0.25 = 0.75$) and the missing column total for "No" ($1 - 0.40 = 0.60$). In a similar way, we can determine the missing number in the "Yes" row ($0.25 - 0.05 = 0.20$) and the "Yes" column ($0.40 - 0.05 = 0.35$). That leaves 0.40 for the "No, No" cell.

	Reads *USA Today*?		
Reads *New York Times*	Yes	No	Total
Yes	0.05	0.20	**0.25**
No	0.35	0.40	**0.75**
Total	**0.40**	**0.60**	**1.00**

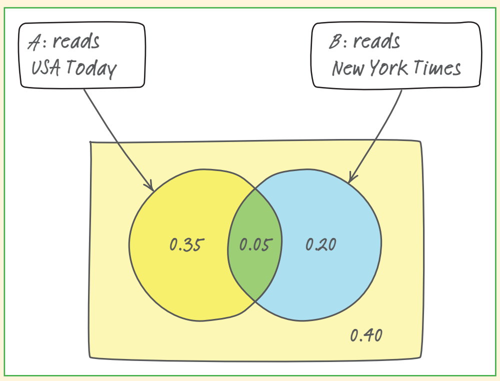


FIGURE 5.7 Venn diagram showing the residents of an apartment complex who A: read *USA Today* and B: read the *New York Times*.

(b) Figure 5.7 shows the Venn diagram that corresponds to the completed two-way table from (a).

(c) If the randomly selected person reads *at least one* of the two papers, then he or she reads *USA Today*, the *New York Times*, or both papers. But that's the same as the event $A \cup B$. From the two-way table, the Venn diagram, or the general addition rule, we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= 0.40 + 0.25 - 0.05 = 0.60$$

So there's a 60% chance that the randomly selected resident reads at least one of the two papers.

(d) From the two-way table or Venn diagram, $P(\text{reads neither paper}) = P(A^C \cap B^C) = 0.40$.

For Practice *Try Exercise* 55

AP® EXAM TIP Many probability problems involve simple computations that you can do on your calculator. It may be tempting to just write down your final answer without showing the supporting work. Don't do it! A "naked answer," even if it's correct, will usually earn you no credit on a free-response question.

In the previous example, the event "reads neither paper" is the complement of the event "reads at least one of the papers." To solve part (d) of the problem, we could have used our answer from (c) and the complement rule:

$$P(\text{reads neither paper}) = 1 - P(\text{reads at least one paper}) = 1 - 0.60 = 0.40$$

As you'll see in Section 5.3, the fact that "none" is the opposite of "at least one" comes in handy for a variety of probability questions.

Section 5.2 Summary

- A **probability model** describes chance behavior by listing the possible outcomes in the **sample space** S and giving the probability that each outcome occurs.
- An **event** is a subset of the possible outcomes in the sample space. To find the probability that an event A happens, we can rely on some basic probability rules:
  - For any event A, $0 \leq P(A) \leq 1$.
  - $P(S) = 1$, where S = the sample space
  - If all outcomes in the sample space are equally likely,

    $$P(A) = \frac{\text{number of outcomes corresponding to event } A}{\text{total number of outcomes in sample space}}$$

  - *Complement rule:* $P(A^C) = 1 - P(A)$, where A^C is the **complement** of event A; that is, the event that A does not happen.
  - *Addition rule for mutually exclusive events:* Events A and B are **mutually exclusive** (**disjoint**) if they have no outcomes in common. If A and B are disjoint, $P(A \text{ or } B) = P(A) + P(B)$.
- A **two-way table** or a **Venn diagram** can be used to display the sample space for a chance process. Two-way tables and Venn diagrams can also be used to find probabilities involving events A and B, like the **union** $(A \cup B)$ and **intersection** $(A \cap B)$. The event $A \cup B$ ("A or B") consists of all outcomes in event A, event B, or both. The event $A \cap B$ ("A and B") consists of outcomes in both A and B.
- The **general addition rule** can be used to find $P(A \text{ or } B)$:

  $$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Section 5.2 Exercises

39. **Role-playing games** Computer games in which the players take the roles of characters are very popular. They go back to earlier tabletop games such as Dungeons & Dragons. These games use many different types of dice. A four-sided die has faces with 1, 2, 3, and 4 spots.

pg 306

(a) List the sample space for rolling the die twice (spots showing on first and second rolls).

(b) What is the assignment of probabilities to outcomes in this sample space? Assume that the die is perfectly balanced.

40. **Tossing coins** Imagine tossing a fair coin 3 times.

(a) What is the sample space for this chance process?

(b) What is the assignment of probabilities to outcomes in this sample space?

41. **Role-playing games** Refer to Exercise 39. Define event A: sum is 5. Find $P(A)$.

42. **Tossing coins** Refer to Exercise 40. Define event B: get more heads than tails. Find $P(B)$.

43. **Probability models?** In each of the following situations, state whether or not the given assignment of probabilities to individual outcomes is legitimate, that is, satisfies the rules of probability. If not, give specific reasons for your answer.

(a) Roll a 6-sided die and record the count of spots on the up-face: $P(1) = 0$, $P(2) = 1/6$, $P(3) = 1/3$, $P(4) = 1/3$, $P(5) = 1/6$, $P(6) = 0$.

(b) Choose a college student at random and record gender and enrollment status: $P(\text{female full-time}) = 0.56$, $P(\text{male full-time}) = 0.44$, $P(\text{female part-time}) = 0.24$, $P(\text{male part-time}) = 0.17$.

(c) Deal a card from a shuffled deck: $P(\text{clubs}) = 12/52$, $P(\text{diamonds}) = 12/52$, $P(\text{hearts}) = 12/52$, $P(\text{spades}) = 16/52$.

44. **Rolling a die** The following figure displays several possible probability models for rolling a die. Some of the models are not *legitimate*. That is, they do not obey the rules. Which are legitimate and which are not? In the case of the illegitimate models, explain what is wrong.

	Probability			
Outcome	Model 1	Model 2	Model 3	Model 4
⚀	1/7	1/3	1/3	1
⚁	1/7	1/6	1/6	1
⚂	1/7	1/6	1/6	2
⚃	1/7	0	1/6	1
⚄	1/7	1/6	1/6	1
⚅	1/7	1/6	1/6	2

45. **Blood types** All human blood can be typed as one of O, A, B, or AB, but the distribution of the types varies a bit with race. Here is the distribution of the blood type of a randomly chosen black American:

pg 308

Blood type:	O	A	B	AB
Probability:	0.49	0.27	0.20	?

(a) What is the probability of type AB blood? Why?

(b) What is the probability that the person chosen does not have type AB blood?

(c) Maria has type B blood. She can safely receive blood transfusions from people with blood types O and B. What is the probability that a randomly chosen black American can donate blood to Maria?

46. **Languages in Canada** Canada has two official languages, English and French. Choose a Canadian at random and ask, "What is your mother tongue?" Here is the distribution of responses, combining many separate languages from the broad Asia/Pacific region:[7]

Language:	English	French	Asian/Pacific	Other
Probability:	0.63	0.22	0.06	?

(a) What probability should replace "?" in the distribution? Why?

(b) What is the probability that a Canadian's mother tongue is not English?

(c) What is the probability that a Canadian's mother tongue is a language other than English or French?

47. **Education among young adults** Choose a young adult (aged 25 to 29) at random. The probability is 0.13 that the person chosen did not complete high school, 0.29 that the person has a high school diploma but no further education, and 0.30 that the person has at least a bachelor's degree.

(a) What must be the probability that a randomly chosen young adult has some education beyond high school but does not have a bachelor's degree? Why?

(b) What is the probability that a randomly chosen young adult has at least a high school education? Which rule of probability did you use to find the answer?

48. **Preparing for the GMAT** A company that offers courses to prepare students for the Graduate Management Admission Test (GMAT) has the following information about its customers: 20% are currently undergraduate students in business; 15% are undergraduate students in other fields of study; 60% are college graduates who are currently employed; and 5% are college graduates who are not employed. Choose a customer at random.

(a) What's the probability that the customer is currently an undergraduate? Which rule of probability did you use to find the answer?

(b) What's the probability that the customer is not an undergraduate business student? Which rule of probability did you use to find the answer?

49. **Who eats breakfast?** Students in an urban school were curious about how many children regularly eat breakfast. They conducted a survey, asking, "Do you

pg 309

eat breakfast on a regular basis?" All 595 students in the school responded to the survey. The resulting data are shown in the two-way table below.[8]

	Male	Female	Total
Eats breakfast regularly	190	110	**300**
Doesn't eat breakfast regularly	130	165	**295**
Total	**320**	**275**	**595**

If we select a student from the school at random, what is the probability that the student is

(a) a female?

(b) someone who eats breakfast regularly?

(c) a female and eats breakfast regularly?

(d) a female or eats breakfast regularly?

50. **Sampling senators** The two-way table below describes the members of the U.S Senate in a recent year.

	Male	Female
Democrats	47	13
Republicans	36	4

If we select a U.S. senator at random, what's the probability that the senator is

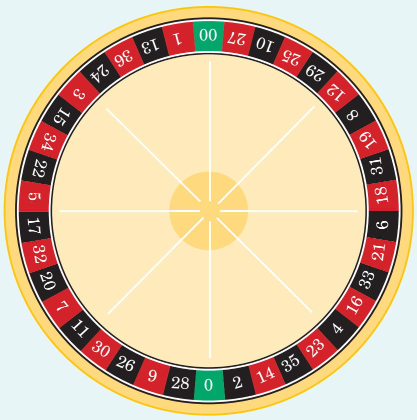
(a) a Democrat?

(b) a female?

(c) a female and a Democrat?

(d) a female or a Democrat?

51. **Roulette** An American roulette wheel has 38 slots with numbers 1 through 36, 0, and 00, as shown in the figure. Of the numbered slots, 18 are red, 18 are black, and 2—the 0 and 00—are green. When the wheel is spun, a metal ball is dropped onto the middle of the wheel. If the wheel is balanced, the ball is equally likely to settle in any of the numbered slots. Imagine spinning a fair wheel once. Define events *B*: ball lands in a black slot, and *E*: ball lands in an even-numbered slot. (Treat 0 and 00 as even numbers.)



(a) Make a two-way table that displays the sample space in terms of events *B* and *E*.

(b) Find $P(B)$ and $P(E)$.

(c) Describe the event "*B* and *E*" in words. Then find $P(B \text{ and } E)$.

(d) Explain why $P(B \text{ or } E) \neq P(B) + P(E)$. Then use the general addition rule to compute $P(B \text{ or } E)$.

52. **Playing cards** Shuffle a standard deck of playing cards and deal one card. Define events *J*: getting a jack, and *R*: getting a red card.

(a) Construct a two-way table that describes the sample space in terms of events *J* and *R*.

(b) Find $P(J)$ and $P(R)$.

(c) Describe the event "*J* and *R*" in words. Then find $P(J \text{ and } R)$.

(d) Explain why $P(J \text{ or } R) \neq P(J) + P(R)$. Then use the general addition rule to compute $P(J \text{ or } R)$.

53. **Who eats breakfast?** Refer to Exercise 49.

(a) Construct a Venn diagram that models the chance process using events *B*: eats breakfast regularly, and *M*: is male.

(b) Find $P(B \cup M)$. Interpret this value in context.

(c) Find $P(B^C \cap M^C)$. Interpret this value in context.

54. **Sampling senators** Refer to Exercise 50.

(a) Construct a Venn diagram that models the chance process using events *R*: is a Republican, and *F*: is female.

(b) Find $P(R \cup F)$. Interpret this value in context.

(c) Find $P(R^C \cap F^C)$. Interpret this value in context.

55. pg 313 **Facebook versus YouTube** A recent survey suggests that 85% of college students have posted a profile on Facebook, 73% use YouTube regularly, and 66% do both. Suppose we select a college student at random.

(a) Make a two-way table for this chance process.

(b) Construct a Venn diagram to represent this setting.

(c) Consider the event that the randomly selected college student has posted a profile on Facebook or uses YouTube regularly. Write this event in symbolic form based on your Venn diagram in part (b).

(d) Find the probability of the event described in part (c). Explain your method.

56. **Mac or PC?** A recent census at a major university revealed that 40% of its students mainly used Macintosh computers (Macs). The rest mainly used PCs. At the time of the census, 67% of the school's students were undergraduates. The rest were graduate students.

In the census, 23% of respondents were graduate students who said that they used PCs as their main computers. Suppose we select a student at random from among those who were part of the census.

(a) Make a two-way table for this chance process.

(b) Construct a Venn diagram to represent this setting.

(c) Consider the event that the randomly selected student is a graduate student and uses a Mac. Write this event in symbolic form based on your Venn diagram in part (b).

(d) Find the probability of the event described in part (c). Explain your method.

Multiple choice: Select the best answer for Exercises 57 to 60.

57. In government data, a household consists of all occupants of a dwelling unit. Choose an American household at random and count the number of people it contains. Here is the assignment of probabilities for the outcome:

Number of persons:	1	2	3	4	5	6	7+
Probability:	0.25	0.32	???	???	0.07	0.03	0.01

The probability of finding 3 people in a household is the same as the probability of finding 4 people. These probabilities are marked ??? in the table of the distribution. The probability that a household contains 3 people must be

(a) 0.68. (b) 0.32. (c) 0.16. (d) 0.08.

(e) between 0 and 1, and we can say no more.

58. In a sample of 275 students, 20 say they are vegetarians. Of the vegetarians, 9 eat both fish and eggs, 3 eat eggs but not fish, and 7 eat neither. Choose one of the vegetarians at random. What is the probability that the chosen student eats fish or eggs?

(a) 9/20 (c) 22/20 (e) 22/275

(b) 13/20 (d) 9/275

Exercises 59 and 60 refer to the following setting. The casino game craps is based on rolling two dice. Here is the assignment of probabilities to the sum of the numbers on the up-faces when two dice are rolled:

Outcome:	2	3	4	5	6	7	8	9	10	11	12
Probability:	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

59. The most common bet in craps is the "pass line." A pass line bettor wins immediately if either a 7 or an 11 comes up on the first roll. This is called a *natural*. What is the probability of a natural?

(a) 2/36 (c) 8/36 (e) 20/36

(b) 6/36 (d) 12/36

60. If a player rolls a 2, 3, or 12, it is called *craps*. What is the probability of getting craps or an even sum on one roll of the dice?

(a) 4/36 (c) 20/36 (e) 32/36

(b) 18/36 (d) 22/36

61. **Crawl before you walk** (3.2) At what age do babies learn to crawl? Does it take longer to learn in the winter, when babies are often bundled in clothes that restrict their movement? Perhaps there might even be an association between babies' crawling age and the average temperature during the month they first try to crawl (around six months after birth). Data were collected from parents who brought their babies to the University of Denver Infant Study Center to participate in one of a number of studies. Parents reported the birth month and the age at which their child was first able to creep or crawl a distance of 4 feet within one minute. Information was obtained on 414 infants (208 boys and 206 girls). Crawling age is given in weeks, and average temperature (in °F) is given for the month that is six months after the birth month.[9]

Birth month	Average crawling age	Average temperature
January	29.84	66
February	30.52	73
March	29.70	72
April	31.84	63
May	28.58	52
June	31.44	39
July	33.64	33
August	32.82	30
September	33.83	33
October	33.35	37
November	33.38	48
December	32.32	57

Analyze the relationship between average crawling age and average temperature. What do you conclude about when babies learn to crawl?

62. **Treating low bone density** (4.2) Fractures of the spine are common and serious among women with advanced osteoporosis (low mineral density in the bones). Can taking strontium ranelate help? A large medical trial assigned 1649 women to take either strontium ranelate or a placebo each day. All of the subjects had osteoporosis and had had at least one fracture. All were taking calcium supplements and receiving standard medical care. The response variables were measurements of bone density and counts of new fractures over three years. The subjects were treated at 10 medical centers in 10 different countries.[10] Outline an appropriate design for this experiment. Explain why this is the proper design.

5.3 Conditional Probability and Independence

WHAT YOU WILL LEARN By the end of the section, you will be able to:

- Calculate and interpret conditional probabilities.
- Use the general multiplication rule to calculate probabilities.
- Use tree diagrams to model a chance process and calculate probabilities involving two or more events.
- Determine if two events are independent.
- When appropriate, use the multiplication rule for independent events to compute probabilities.

The probability we assign to an event can change if we know that some other event has occurred. This idea is the key to many applications of probability.

What Is Conditional Probability?

Let's return to the setting of the pierced-ears example (page 309). Earlier, we used the two-way table below to find probabilities involving events A: is male and B: has pierced ears for a randomly selected student. Here is a summary of our previous results:

$P(A) = P(\text{male}) = 90/178$ $\quad P(A \cap B) = P(\text{male and pierced ears}) = 19/178$

$P(B) = P(\text{pierced ears}) = 103/178$ $\quad P(A \cup B) = P(\text{male or pierced ears}) = 174/178$

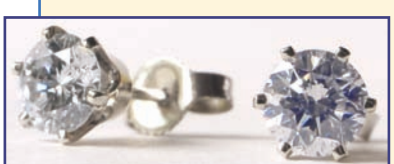
	Gender		
Pierced Ears	Male	Female	Total
Yes	19	84	**103**
No	71	4	**75**
Total	**90**	**88**	**178**

Now let's turn our attention to some other interesting probability questions.

EXAMPLE

Who Has Pierced Ears?

The idea of conditional probability



1. If we know that a randomly selected student has pierced ears, what is the probability that the student is male? There are a total of 103 students in the class with pierced ears. We can restrict our attention to this group, since we are told that the chosen student has pierced ears. Because there are 19 males among the 103 students with pierced ears, the desired probability is

$$P(\text{is male } \textit{given} \text{ has pierced ears}) = 19/103, \text{ or about } 18.4\%$$

2. If we know that a randomly selected student is male, what's the probability that the student has pierced ears? This time, our attention is focused on the males in the class. Because 19 of the 90 males in the class have pierced ears,

$$P(\text{has pierced ears } \textit{given} \text{ is male}) = 19/90, \text{ or about } 21.1\%$$

These two questions sound alike, but they're asking about two very different things.

A probability like "the probability that a randomly selected student is male *given that* the student has pierced ears" is known as a **conditional probability**. The name comes from the fact that we are trying to find the probability that one event will happen under the *condition* that some other event is already known to have occurred. We often use the phrase "given that" to signal the condition. There's even a special notation to indicate a conditional probability. In the example above, we would write $P(\text{is male} \mid \text{has pierced ears})$, where the | means "given that" or "under the condition that." Because we already defined the events A: is male and B: has pierced ears, we could write the conditional probability as $P(A \mid B) = 19/103$.

DEFINITION: Conditional probability

The probability that one event happens given that another event is already known to have happened is called a **conditional probability**. Suppose we know that event *A* has happened. Then the probability that event *B* happens given that event *A* has happened is denoted by $P(B \mid A)$.

Let's look more closely at how conditional probabilities are calculated. From the two-way table below, we see that

$$P(\text{male} \mid \text{pierced ears}) = \frac{\text{Number of students who are male and have pierced ears}}{\text{Number of students with pierced ears}} = \frac{19}{103}$$

	Gender			
Pierced Ears?	Male	Female	Total	
Yes	19	84	**103**	$P(\text{male} \mid \text{pierced ears}) = 19/103$
No	71	4	**75**	
Total	**90**	**88**	**178**	

What if we focus on probabilities instead of numbers of students? Notice that

$$\frac{P(\text{male and pierced ears})}{P(\text{pierced ears})} = \frac{\frac{19}{178}}{\frac{103}{178}} = \frac{19}{103} = P(\text{male} \mid \text{pierced ears})$$

In symbols, $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$. This observation leads to a general formula for computing a conditional probability.

CALCULATING CONDITIONAL PROBABILITIES

To find the conditional probability $P(A \mid B)$, use the formula

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

The conditional probability $P(B \mid A)$ is given by

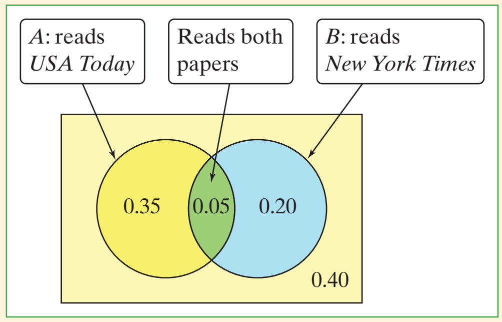
$$P(B \mid A) = \frac{P(B \cap A)}{P(A)}$$

Here's an example that illustrates the use of this formula in a familiar setting.

EXAMPLE Who Reads the Paper?

Conditional probability formula

On page 313, we classified the residents of a large apartment complex based on the events A: reads *USA Today*, and B: reads the *New York Times*. The completed Venn diagram is reproduced here.



AP® EXAM TIP You can write statements like $P(B \mid A)$ if events A and B are defined clearly, or you can use a verbal equivalent, such as P(reads *New York Times* | reads *USA Today*). Use the approach that makes the most sense to you.

PROBLEM: What's the probability that a randomly selected resident who reads *USA Today* also reads the *New York Times*?

SOLUTION: Because we're given that the randomly chosen resident reads *USA Today*, we want to find P(reads *New York Times* | reads *USA Today*), or $P(B \mid A)$. By the conditional probability formula,

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)}$$

Because $P(B \cap A) = 0.05$ and $P(A) = 0.40$, we have

$$P(B \mid A) = \frac{0.05}{0.40} = 0.125$$

There's a 12.5% chance that a randomly selected resident who reads *USA Today* also reads the *New York Times*.

For Practice *Try Exercise* **71**

THINK ABOUT IT

Is there a connection between conditional probability and the conditional distributions of Chapter 1? Of course! For the college statistics class that we discussed earlier, Figure 5.8 shows the conditional distribution of ear-piercing status for each gender. Above, we found that P(pierced ears | male) = 19/90, or about 21%. Note that P(pierced ears | female) = 84/88, or about 95%. You can see these values displayed in the "Yes" bars of Figure 5.8.

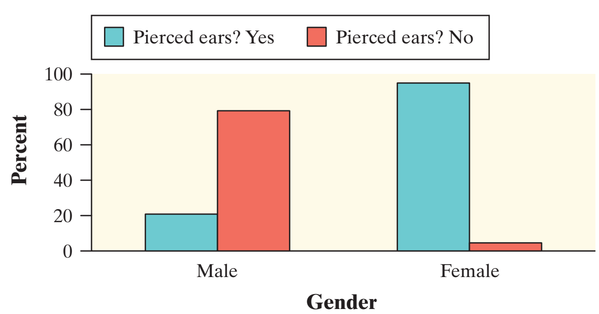


FIGURE 5.8 Conditional distribution of ear-piercing status for each gender in a large college statistics class.

CHECK YOUR UNDERSTANDING

Students at the University of New Harmony received 10,000 course grades last semester. The two-way table below breaks down these grades by which school of the university taught the course. The schools are Liberal Arts, Engineering and Physical Sciences, and Health and Human Services.

	Grade Level		
School	**A**	**B**	**Below B**
Liberal Arts	2,142	1,890	2,268
Engineering and Physical Sciences	368	432	800
Health and Human Services	882	630	588

(This table is based closely on grade distributions at an actual university, simplified a bit for clarity.)[11]

College grades tend to be lower in engineering and the physical sciences (EPS) than in liberal arts and social sciences (which includes Health and Human Services). Choose a University of New Harmony course grade at random. Consider the two events E: the grade comes from an EPS course, and L: the grade is lower than a B.

1. Find $P(L)$. Interpret this probability in context.
2. Find $P(E \mid L)$ and $P(L \mid E)$. Which of these conditional probabilities tells you whether this college's EPS students tend to earn lower grades than students in liberal arts and social sciences? Explain.

The General Multiplication Rule and Tree Diagrams

Suppose that A and B are two events resulting from the same chance process. We can find the probability $P(A \text{ or } B)$ with the general addition rule:

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

How do we find the probability that both events happen, $P(A \text{ and } B)$?

Start with the conditional probability formula

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)}$$

The numerator, $P(B \cap A)$, is the probability we want because $P(B \text{ and } A)$ is the same as $P(A \text{ and } B)$. Multiply both sides of the above equation by $P(A)$ to get

$$P(A) \cdot P(B \mid A) = P(B \cap A) = P(A \cap B) = P(A \text{ and } B)$$

This formula is known as the **general multiplication rule**.

GENERAL MULTIPLICATION RULE

The probability that events A and B both occur can be found using the **general multiplication rule**

$$P(A \text{ and } B) = P(A \cap B) = P(A) \cdot P(B \mid A)$$

where $P(B \mid A)$ is the conditional probability that event B occurs given that event A has already occurred.

In words, this rule says that for both of two events to occur, first one must occur, and then given that the first event has occurred, the second must occur. This is just common sense expressed in the language of probability, as the following example illustrates.

EXAMPLE Teens with Online Profiles

Using the general multiplication rule

The Pew Internet and American Life Project find that 93% of teenagers (ages 12 to 17) use the Internet, and that 55% of online teens have posted a profile on a social-networking site.[12]

PROBLEM: Find the probability that a randomly selected teen uses the Internet and has posted a profile. Show your work.

SOLUTION: We know that $P(\text{online}) = 0.93$ and $P(\text{profile} \mid \text{online}) = 0.55$. Use the general multiplication rule:

$$P(\text{online and have profile}) = P(\text{online}) \cdot P(\text{profile} \mid \text{online}) = (0.93)(0.55) = 0.5115$$

There is about a 51% chance that a randomly selected teen uses the Internet and has posted a profile on a social-networking site.

For Practice *Try Exercise* 73

The general multiplication rule is especially useful when a chance process involves a sequence of outcomes. In such cases, we can use a **tree diagram** to display the sample space.

EXAMPLE Serve It Up!

Tree diagrams and the general multiplication rule

Tennis great Roger Federer made 63% of his first serves in the 2011 season. When Federer made his first serve, he won 78% of the points. When Federer missed his first serve and had to serve again, he won only 57% of the points.[13] Suppose we randomly choose a point on which Federer served.

Figure 5.9 on the facing page shows a tree diagram for this chance process. There are only two possible outcomes on Federer's first serve, a make or a miss. The first set of branches in the tree diagram displays these outcomes with their probabilities. The second set of branches shows the two possible results of the point for Federer—win or lose—and the chance of each result based on the outcome of the first serve. Note that the probabilities on the second set of branches are *conditional probabilities*, like $P(\text{win point} \mid \text{make first serve}) = 0.78$.

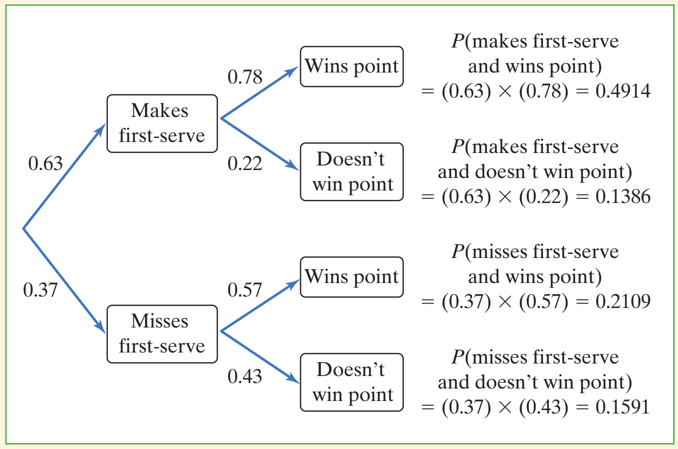


FIGURE 5.9 A tree diagram displaying the sample space of randomly choosing a point on which Roger Federer served.

What's the probability that Federer makes the first serve and wins the point? From the tree diagram, Federer makes 63% of his first serves. Of this 63%, he wins the point 78% of the time. Because 78% of 63% = (0.63)(0.78) = 0.4914, Federer makes his first serve and wins the point about 49.14% of the time.

The previous calculation amounts to multiplying probabilities along the branches of the tree diagram. Why does this work? The general multiplication rule provides the answer:

$$\begin{aligned} P(\text{make first serve and win point}) &= P(\text{make first serve}) \cdot P(\text{win point} \mid \text{make first serve}) \\ &= (0.63)(0.78) = 0.4914 \end{aligned}$$

When Federer is serving, what's the probability that he wins the point? From the tree diagram, there are two ways Federer can win the point. He can make the first serve and win the point, or he can miss the first serve and win the point. Because these outcomes are mutually exclusive,

$$\begin{aligned} P(\text{win point}) &= P(\text{makes first serve and wins point}) + P(\text{misses first serve and wins point}) \\ &= 0.4914 + 0.2109 = 0.7023 \end{aligned}$$

Federer wins about 70% of the points when he is serving.

Some interesting conditional probability questions involve "going in reverse" on a tree diagram. Here's one related to the previous example. Suppose you are watching a recording of one of Federer's matches from 2011 and he is serving in the current game. You get distracted before seeing his first serve but look up in time to see Federer win the point. How likely is it that he missed his first serve?

To find this probability, we start with the result of the point, which is displayed on the second set of branches in Figure 5.9, and ask about the outcome of the serve, which is shown on the first set of branches. We can use the information from the tree diagram and the conditional probability to do the required calculation:

Some people use a result known as Bayes's theorem to solve probability questions that require "going backward" in a tree diagram, like the one in this example. To be honest, Bayes's theorem is just a complicated formula for computing conditional probabilities. For that reason, we won't introduce it.

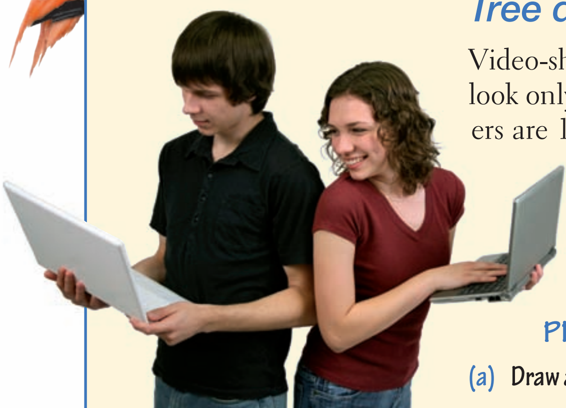
$$\begin{aligned} P(\text{missed first serve} \mid \text{wins point}) &= \frac{P(\text{missed first serve and wins point})}{P(\text{wins point})} \\ &= \frac{0.2109}{0.4914 + 0.2109} = \frac{0.2109}{0.7023} = 0.3003 \end{aligned}$$

Given that Federer won the point, there is about a 30% chance that he missed his first serve.

Here is another example where we need to reverse the conditioning.

EXAMPLE Who Visits YouTube?

Tree diagrams and conditional probability



Video-sharing sites, led by YouTube, are popular destinations on the Internet. Let's look only at adult Internet users, aged 18 and over. About 27% of adult Internet users are 18 to 29 years old, another 45% are 30 to 49 years old, and the remaining 28% are 50 and over. The Pew Internet and American Life Project finds that 70% of Internet users aged 18 to 29 have visited a video-sharing site, along with 51% of those aged 30 to 49 and 26% of those 50 or older. Do most Internet users visit YouTube and similar sites?

PROBLEM: Suppose we select an adult Internet user at random.

(a) Draw a tree diagram to represent this situation.

(b) Find the probability that this person has visited a video-sharing site. Show your work.

(c) Given that this person has visited a video-sharing site, find the probability that he or she is aged 18 to 29. Show your work.

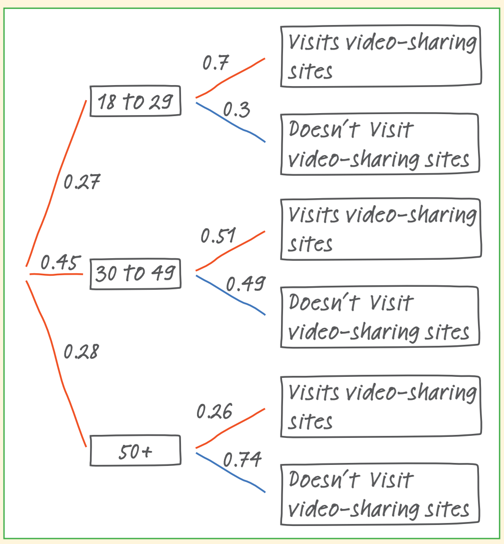


FIGURE 5.10 Tree diagram for use of the Internet and video-sharing sites such as YouTube. The three disjoint paths to the outcome that an adult Internet user visits video-sharing sites are colored red.

SOLUTION:

(a) The tree diagram in Figure 5.10 organizes the given information.

(b) There are three disjoint paths to "visits video-sharing sites," one for each of the three age groups. These paths are colored red in Figure 5.10. Because the three paths are disjoint, the probability that an adult Internet user visits video-sharing sites is the sum of their probabilities:

$$\begin{aligned} P(\text{visits video-sharing sites}) &= (0.27)(0.70) + (0.45)(0.51) + (0.28)(0.26) \\ &= 0.1890 + 0.2295 + 0.0728 = 0.4913 \end{aligned}$$

About 49% of all adult Internet users have visited a video-sharing site.

(c) Use the tree diagram and the definition of conditional probability:

$$\begin{aligned} P(\text{18 to 29} \mid \text{visits video-sharing site}) &= \frac{P(\text{18 to 29 and visits video-sharing site})}{P(\text{visits video-sharing site})} \\ &= \frac{0.1890}{0.4913} = 0.3847 \end{aligned}$$

Given that the person visits video-sharing sites, there is about a 38.5% chance that he or she is aged 18 to 29.

For Practice *Try Exercise* 77

One of the most important applications of tree diagrams and conditional probability is in the area of drug and disease testing.

EXAMPLE Mammograms

Conditional probability in real life

Many women choose to have annual mammograms to screen for breast cancer after age 40. A mammogram isn't foolproof. Sometimes, the test suggests that a woman has breast cancer when she really doesn't (a "false positive"). Other times, the test says that a woman doesn't have breast cancer when she actually does (a "false negative").

Suppose that we know the following information about breast cancer and mammograms in a particular region:

- One percent of the women aged 40 or over in this region have breast cancer.
- For women who have breast cancer, the probability of a negative mammogram is 0.03.
- For women who don't have breast cancer, the probability of a positive mammogram is 0.06.

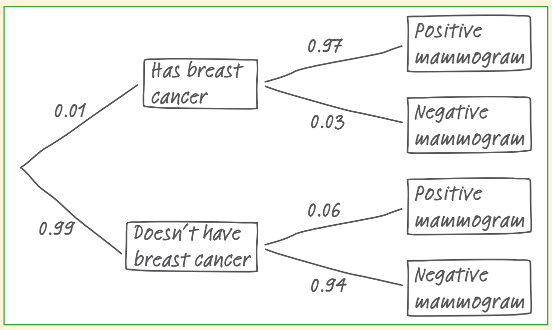


FIGURE 5.11 Tree diagram showing whether or not a woman has breast cancer and the likelihood of her receiving a positive or a negative test result from a mammogram.

PROBLEM: A randomly selected woman aged 40 or over from this region tests positive for breast cancer in a mammogram. Find the probability that she actually has breast cancer. Show your work.

SOLUTION: The tree diagram in Figure 5.11 summarizes the situation. Because 1% of women in this region have breast cancer, 99% don't. Of those women who do have breast cancer, 3% would test negative on a mammogram. The remaining 97% would (correctly) test positive. Among the women who don't have breast cancer, 6% would test positive on a mammogram. The remaining 94% would (correctly) test negative.

We want to find $P(\text{breast cancer} \mid \text{positive mammogram})$. By the conditional probability formula,

$$P(\text{breast cancer} \mid \text{positive mammogram}) = \frac{P(\text{breast cancer and positive mammogram})}{P(\text{positive mammogram})}$$

- To find $P(\text{breast cancer and positive mammogram})$, we use the general multiplication rule along with the information displayed in the tree diagram:

$$\begin{aligned} P(\text{breast cancer and positive mammogram}) &= P(\text{breast cancer}) \cdot P(\text{positive mammogram} \mid \text{breast cancer}) \\ &= (0.01)(0.97) = 0.0097 \end{aligned}$$

- To find $P(\text{positive mammogram})$, we need to calculate the probability that a randomly selected woman aged 40 or over from this region gets a positive mammogram. There are two ways this can happen: (1) if the woman has breast cancer and the test result is positive, and (2) if the woman doesn't have cancer, but the mammogram gives a false positive. From the tree diagram, the desired probability is

$$P(\text{positive mammogram}) = (0.01)(0.97) + (0.99)(0.06) = 0.0691$$

Using these two results, we can find the conditional probability:

$$P(\text{breast cancer} \mid \text{positive mammogram}) = \frac{P(\text{breast cancer and positive mammogram})}{P(\text{positive mammogram})} = \frac{0.0097}{0.0691} = 0.14$$

Given that a randomly selected woman from the region has a positive mammogram, there is only about a 14% chance that she actually has breast cancer!

For Practice *Try Exercise* 81

Are you surprised by the final result of the example? Most people are. Sometimes a two-way table that includes counts is more convincing. To make calculations simple, we'll suppose that there were exactly 10,000 women aged 40 or over in this region, and that exactly 100 have breast cancer (that's 1% of the women).

How many of those 100 would have a positive mammogram? It would be 97% of 100, or 97 of them. That leaves 3 who would test negative. How many of the 9900 women who don't have breast cancer would get a positive mammogram? Six percent of them, or $(0.06)(9900) = 594$ women. The remaining $9900 - 594 = 9306$ would test negative. In total, $97 + 594 = 691$ women would have positive mammograms and $3 + 9306 = 9309$ women would have negative mammograms. This information is summarized in the two-way table below:

		Has breast cancer?		
		Yes	No	Total
Mammogram result	Positive	97	594	**691**
	Negative	3	9306	**9309**
	Total	**100**	**9900**	**10,000**

Given that a randomly selected woman has a positive mammogram, the two-way table shows that the conditional probability $P(\text{breast cancer} \mid \text{positive mammogram}) = 97/691 = 0.14$.

This example illustrates an important fact when considering proposals for widespread testing for serious diseases or illegal drug use: if the condition being tested is uncommon in the population, many positives will be false positives. The best remedy is to retest any individual who tests positive.

CHECK YOUR UNDERSTANDING

A computer company makes desktop and laptop computers at factories in three states—California, Texas, and New York. The California factory produces 40% of the company's computers, the Texas factory makes 25%, and the remaining 35% are manufactured in New York. Of the computers made in California, 75% are laptops. Of those made in Texas and New York, 70% and 50%, respectively, are laptops. All computers are first shipped to a distribution center in Missouri before being sent out to stores. Suppose we select a computer at random from the distribution center.[14]

1. Construct a tree diagram to represent this situation.
2. Find the probability that the computer is a laptop. Show your work.
3. Given that a laptop is selected, what is the probability that it was made in California?

Conditional Probability and Independence

Suppose you toss a fair coin twice. Define events A: first toss is a head, and B: second toss is a head. We know that $P(A) = 1/2$ and $P(B) = 1/2$. What's $P(B \mid A)$? It's the conditional probability that the second toss is a head given that the first toss was a head. The coin has no memory, so $P(B \mid A) = 1/2$. In this case, $P(B \mid A) = P(B)$. Knowing that the first toss was a head does not change the probability that the second toss is a head.

Let's contrast the coin-toss scenario with our earlier pierced-ears example. In that case, the chance process involved randomly selecting a student from a college statistics class. The events of interest were A: is male, and B: has pierced ears. We already found that $P(A) = 90/178$, $P(B) = 103/178$, and $P(B \mid A) = 19/90$. If we know that the chosen student is male, the probability that he has pierced ears is $19/90 = 0.211$. This conditional probability is very different from the unconditional probability $P(B) = 103/178 = 0.579$ that a randomly selected student from the class has pierced ears.

	Gender		
Pierced Ears	Male	Female	Total
Yes	19	84	**103**
No	71	4	**75**
Total	**90**	**88**	**178**

To recap, $P(B \mid A) = P(B)$ for events A and B in the coin-toss setting. For the pierced-ears scenario, however, $P(B \mid A) \neq P(B)$. When knowledge that one event has happened does not change the likelihood that another event will happen, we say that the two events are **independent.**

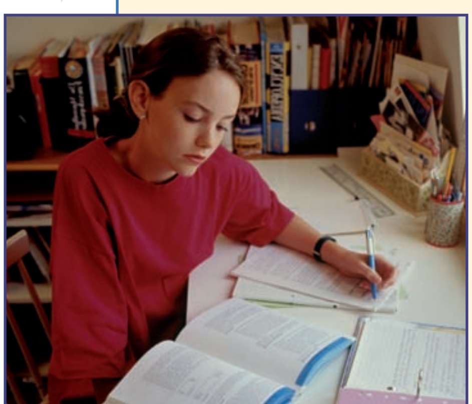
DEFINITION: Independent events

Two events A and B are **independent** if the occurrence of one event does not change the probability that the other event will happen. In other words, events A and B are independent if $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$.

Determining whether two events related to the same chance process are independent requires us to compute probabilities. Here's an example that shows what we mean.

EXAMPLE Lefties Down Under

Checking for independence



Is there a relationship between gender and handedness? To find out, we used CensusAtSchool's Random Data Selector to choose an SRS of 100 Australian high school students who completed a survey. The two-way table displays data on the gender and dominant hand of each student.

	Gender		
Dominant Hand	Male	Female	Total
Right	39	51	**90**
Left	7	3	**10**
Total	**46**	**54**	**100**

PROBLEM: Are the events "male" and "left-handed" independent? Justify your answer.

SOLUTION: To check whether the two events are independent, we want to find out if knowing that one event has happened changes the probability that the other event occurs. Suppose we are told that the chosen student is male. From the two-way table, $P(\text{left-handed} \mid \text{male}) = 7/46 = 0.152$. The unconditional probability $P(\text{left-handed}) = 10/100 = 0.10$. These two probabilities are close, but they're not equal. So the events "male" and "left-handed" are not independent. Knowing that the student is male increases the probability that the student is left-handed.

For Practice *Try Exercise* 85

You might have thought, "Surely there's no connection between gender and handedness. The events 'male' and 'left-handed' are bound to be independent." As the example shows, you can't use your intuition to check whether events are independent. To be sure, you have to calculate some probabilities.

In the preceding example, we could have also compared $P(\text{male} \mid \text{left-handed})$ with $P(\text{male})$. Of the 10 left-handed students in the sample, 7 were male. So $P(\text{male} \mid \text{left-handed}) = 7/10 = 0.70$. We can see from the two-way table that $P(\text{male}) = 46/100 = 0.46$. Once again, the two probabilities are not equal. Knowing that a person is left-handed makes it more likely that the person is male.

THINK ABOUT IT

Is there a connection between independence of events and association between two variables? In the previous example, we found that the events "male" and "left-handed" were *not* independent. Does that mean there actually *is* a relationship between the variables gender and handedness in the larger population? Maybe or maybe not. If there is *no* association between the variables, it would be surprising to choose a random sample of 100 students for which $P(\text{left-handed} \mid \text{male})$ was *exactly* equal to $P(\text{left-handed})$. But these two probabilities should be close to equal if there's no association between the variables. How close is close? You'll have to wait a few chapters to find out.

CHECK YOUR UNDERSTANDING

For each chance process below, determine whether the events are independent. Justify your answer.

1. Shuffle a standard deck of cards, and turn over the top card. Put it back in the deck, shuffle again, and turn over the top card. Define events A: first card is a heart, and *B*: second card is a heart.

2. Shuffle a standard deck of cards, and turn over the top two cards, one at a time. Define events A: first card is a heart, and *B*: second card is a heart.

3. The 28 students in Mr. Tabor's AP® Statistics class completed a brief survey. One of the questions asked whether each student was right- or left-handed. The two-way table summarizes the class data. Choose a student from the class at random. The events of interest are "female" and "right-handed."

	Gender	
Handedness	**Female**	**Male**
Left	3	1
Right	18	6

Independence: A Special Multiplication Rule

What happens to the general multiplication rule in the special case when events A and *B* are independent? In that case, $P(B \mid A) = P(B)$. We can simplify the general multiplication rule as follows:

$$\begin{aligned} P(A \cap B) &= P(A) \cdot P(B \mid A) \\ &= P(A) \cdot P(B) \end{aligned}$$

This result is known as the **multiplication rule for independent events.**

DEFINITION: Multiplication rule for independent events

If *A* and *B* are independent events, then the probability that *A* and *B* both occur is

$$P(A \cap B) = P(A) \cdot P(B)$$

Note that this rule *only* applies to independent events. Let's look at an example that uses the multiplication rule for independent events to analyze an important historical event.

EXAMPLE

The *Challenger* Disaster

Independence and the multiplication rule

On January 28, 1986, Space Shuttle *Challenger* exploded on takeoff. All seven crew members were killed. Following the disaster, scientists and statisticians helped analyze what went wrong. They determined that the failure of O-ring joints in the shuttle's booster rockets was to blame. Under the cold conditions that day, experts estimated that the probability that an individual O-ring joint would function properly was 0.977. But there were six of these O-ring joints, and all six had to function properly for the shuttle to launch safely.

PROBLEM: Assuming that O-ring joints succeed or fail independently, find the probability that the shuttle would launch safely under similar conditions.

SOLUTION: For the shuttle to launch safely, all six O-ring joints need to function properly. The chance that this happens is given by

$$P(\text{joint 1 OK } \textit{and} \text{ joint 2 OK } \textit{and} \text{ joint 3 OK } \textit{and} \text{ joint 4 OK } \textit{and} \text{ joint 5 OK } \textit{and} \text{ joint 6 OK})$$

By the multiplication rule for independent events, this probability is

$$P(\text{joint 1 OK}) \cdot P(\text{joint 2 OK}) \cdot P(\text{joint 3 OK}) \cdot P(\text{joint 4 OK}) \cdot P(\text{joint 5 OK}) \cdot P(\text{joint 6 OK})$$
$$= (0.977)(0.977)(0.977)(0.977)(0.977)(0.977) = 0.87$$

There's an 87% chance that the shuttle would launch safely under similar conditions (and a 13% chance that it wouldn't).

Note: As a result of the statistical analysis following the *Challenger* disaster, NASA made important safety changes to the design of the shuttle's booster rockets.

For Practice *Try Exercise* 89

The next example uses the fact that "at least one" and "none" are opposites.

EXAMPLE

Rapid HIV Testing

Finding the probability of "at least one"

Many people who come to clinics to be tested for HIV, the virus that causes AIDS, don't come back to learn the test results. Clinics now use "rapid HIV tests" that give a result while the client waits. In a clinic in Malawi, for example,

use of rapid tests increased the percent of clients who learned their test results from 69% to 99.7%.

The trade-off for fast results is that rapid tests are less accurate than slower laboratory tests. Applied to people who have no HIV antibodies, one rapid test has probability about 0.004 of producing a false positive (that is, of falsely indicating that antibodies are present).[15]

PROBLEM: If a clinic tests 200 randomly selected people who are free of HIV antibodies, what is the chance that at least one false positive will occur?

SOLUTION: It is reasonable to assume that the test results for different individuals are independent. We have 200 independent events, each with probability 0.004. "At least one" combines many possible outcomes. It will be easier to use the fact that

$$P(\text{at least one positive}) = 1 - P(\text{no positives})$$

We'll find P(no positives) first. The probability of a negative result for any one person is $1 - 0.004 = 0.996$. To find the probability that all 200 people tested have negative results, use the multiplication rule for independent events:

$$\begin{aligned} P(\text{no positives}) &= P(\text{all 200 negative}) \\ &= (0.996)(0.996)\ldots(0.996) \\ &= 0.996^{200} = 0.4486 \end{aligned}$$

The probability we want is therefore

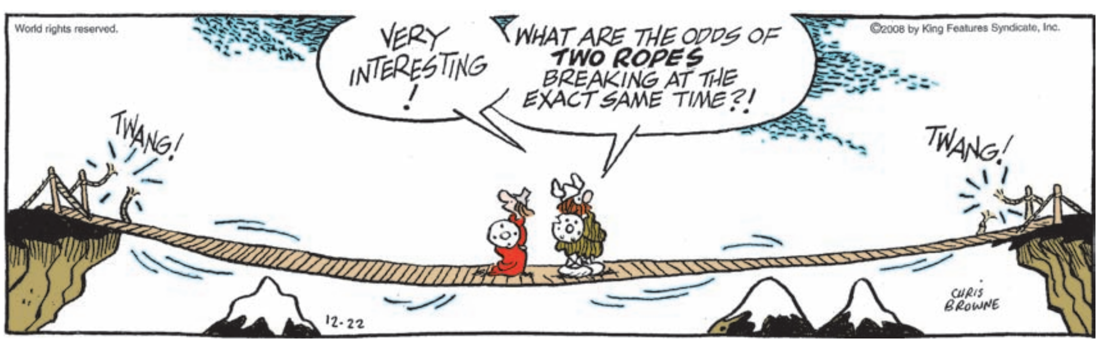
$$P(\text{at least one positive}) = 1 - 0.4486 = 0.5514$$

There is more than a 50% chance that at least 1 of the 200 people will test positive for HIV, even though no one has the virus.

For Practice *Try Exercise* **91**

The multiplication rule $P(A \text{ and } B) = P(A) \cdot P(B)$ holds if A and B are *independent* but not otherwise. The addition rule $P(A \text{ or } B) = P(A) + P(B)$ holds if A and B are *mutually exclusive* but not otherwise. Resist the temptation to use these simple rules when the conditions that justify them are not present.

caution



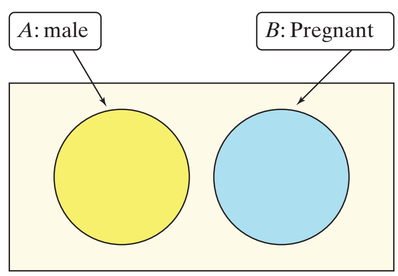
EXAMPLE Sudden Infant Death

Condemned by independence

Assuming independence when it isn't true can lead to disaster. Several mothers in England were convicted of murder simply because two of their children had died in their cribs with no visible cause. An "expert witness" for the prosecution said that the probability of an unexplained crib death in a nonsmoking middle-class family is 1/8500. He then multiplied 1/8500 by 1/8500 to claim that there is only a 1-in-72-million chance that two children in the same family could have died naturally. This is nonsense: it assumes that crib deaths are independent, and data suggest that they are not. Some common genetic or environmental cause, not murder, probably explains the deaths.

THINK ABOUT IT

Is there a connection between mutually exclusive and independent? Let's start with a new chance process. Choose a U.S. adult at random. Define event A: the person is male, and event B: the person is pregnant. It's fairly clear that these two events are mutually exclusive (can't happen together)! What about independence? If you know that event A has occurred, does this change the probability that event B happens? Of course! If we know the person is male, then the chance that the person is pregnant is 0. Because $P(B \mid A) \neq P(B)$, the two events are not independent. *Two mutually exclusive events can never be independent*, because if one event happens, the other event is guaranteed not to happen.



CHECK YOUR UNDERSTANDING

1. During World War II, the British found that the probability that a bomber is lost through enemy action on a mission over occupied Europe was 0.05. Assuming that missions are independent, find the probability that a bomber returned safely from 20 missions.

2. Government data show that 8% of adults are full-time college students and that 30% of adults are age 55 or older. Because $(0.08)(0.30) = 0.024$, can we conclude that about 2.4% of adults are college students 55 or older? Why or why not?

case closed

Calculated Risks

The chapter-opening Case Study on page 287 described drug-testing programs for high school athletes. Suppose that 16% of the high school athletes in a large school district have taken a banned substance. The drug test used by this district has a false positive rate of 5% and a false negative rate of 10%. Use what you have learned in this chapter to help answer the following questions about the district's drug-testing program. Show your method clearly.

1. What's the probability that a randomly chosen athlete tests positive for banned substances?
2. If two athletes are randomly selected, what's the probability that at least one of them tests positive?
3. If a randomly chosen athlete tests positive, what's the probability that the student did not take a banned substance? Based on your answer, do you think that an athlete who tests positive should be suspended from athletic competition for a year? Why or why not?
4. If a randomly chosen athlete tests negative, what's the probability that the student took a banned substance? Explain why it makes sense for the drug-testing process to be designed so that this probability is less than the one you found in Question 3.
5. The district decides to immediately retest any athlete who tests positive. Assume that the results of an athlete's two tests are independent. Find the probability that a student who gets a positive result on both tests actually took a banned substance. Based on your answer, do you think that an athlete who tests positive twice should be suspended from athletic competition for a year? Why or why not?

Section 5.3 Summary

- If one event has happened, the chance that another event will happen is a **conditional probability**. The notation $P(B \mid A)$ represents the probability that event B occurs given that event A has occurred.
- You can calculate conditional probabilities with the **conditional probability formula**

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- The **general multiplication rule** states that the probability of events A and B occurring together is

$$P(A \text{ and } B) = P(A \cap B) = P(A) \cdot P(B \mid A)$$

- When chance behavior involves a sequence of outcomes, a **tree diagram** can be used to describe the sample space. Tree diagrams can also help in finding the probability that two or more events occur together. We simply multiply along the branches that correspond to the outcomes of interest.
- When knowing that one event has occurred does not change the probability that another event happens, we say that the two events are **independent**. For independent events A and B, $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$. If two events A and B are mutually exclusive (disjoint), they cannot be independent.
- In the special case of *independent* events, the multiplication rule becomes

$$P(A \text{ and } B) = P(A \cap B) = P(A) \cdot P(B)$$

Section 5.3 Exercises

63. **Get rich** A survey of 4826 randomly selected young adults (aged 19 to 25) asked, "What do you think are the chances you will have much more than a middle-class income at age 30?" The two-way table shows the responses.[16] Choose a survey respondent at random.

	Gender		
Opinion	**Female**	**Male**	**Total**
Almost no chance	96	98	**194**
Some chance but probably not	426	286	**712**
A 50-50 chance	696	720	**1416**
A good chance	663	758	**1421**
Almost certain	486	597	**1083**
Total	**2367**	**2459**	**4826**

(a) Given that the person selected is male, what's the probability that he answered "almost certain"?

(b) If the person selected said "some chance but probably not," what's the probability that the person is female?

64. **A *Titanic* disaster** In 1912 the luxury liner *Titanic*, on its first voyage across the Atlantic, struck an iceberg and sank. Some passengers got off the ship in lifeboats, but many died. The two-way table gives information about adult passengers who lived and who died, by class of travel. Suppose we choose an adult passenger at random.

	Survival Status	
Class of Travel	**Survived**	**Died**
First class	197	122
Second class	94	167
Third class	151	476

(a) Given that the person selected was in first class, what's the probability that he or she survived?

(b) If the person selected survived, what's the probability that he or she was a third-class passenger?

65. **Sampling senators** The two-way table describes the members of the U.S. Senate in a recent year. Suppose we select a senator at random. Consider events D: is a democrat, and F: is female.

	Male	**Female**
Democrats	47	13
Republicans	36	4

(a) Find $P(D \mid F)$. Explain what this value means.

(b) Find $P(F \mid D)$. Explain what this value means.

66. **Who eats breakfast?** The following two-way table describes the 595 students who responded to a school survey about eating breakfast. Suppose we select a student at random. Consider events B: eats breakfast regularly, and M: is male.

	Male	Female	Total
Eats breakfast regularly	190	110	**300**
Doesn't eat breakfast regularly	130	165	**295**
Total	**320**	**275**	**595**

(a) Find $P(B \mid M)$. Explain what this value means.

(b) Find $P(M \mid B)$. Explain what this value means.

67. **Foreign-language study** Choose a student in grades 9 to 12 at random and ask if he or she is studying a language other than English. Here is the distribution of results:

Language:	Spanish	French	German	All others	None
Probability:	0.26	0.09	0.03	0.03	0.59

(a) What's the probability that the student is studying a language other than English?

(b) What is the conditional probability that a student is studying Spanish given that he or she is studying some language other than English?

68. **Income tax returns** Here is the distribution of the adjusted gross income (in thousands of dollars) reported on individual federal income tax returns in a recent year:

Income:	<25	25–49	50–99	100–499	≥ 500
Probability:	0.431	0.248	0.215	0.100	0.006

(a) What is the probability that a randomly chosen return shows an adjusted gross income of \$50,000 or more?

(b) Given that a return shows an income of at least \$50,000, what is the conditional probability that the income is at least \$100,000?

69. **Tall people and basketball players** Select an adult at random. Define events T: person is over 6 feet tall, and B: person is a professional basketball player. Rank the following probabilities from smallest to largest. Justify your answer.

$P(T)$ $P(B)$ $P(T \mid B)$ $P(B \mid T)$

70. **Teachers and college degrees** Select an adult at random. Define events A: person has earned a college degree, and T: person's career is teaching. Rank the following probabilities from smallest to largest. Justify your answer.

$P(A)$ $P(T)$ $P(A \mid T)$ $P(T \mid A)$

71. **Facebook versus YouTube** A recent survey suggests that 85% of college students have posted a profile on Facebook, 73% use YouTube regularly, and 66% do both. Suppose we select a college student at random and learn that the student has a profile on Facebook. Find the probability that the student uses YouTube regularly. Show your work.

pg 320

72. **Mac or PC?** A recent census at a major university revealed that 40% of its students mainly used Macintosh computers (Macs). The rest mainly used PCs. At the time of the census, 67% of the school's students were undergraduates. The rest were graduate students. In the census, 23% of the respondents were graduate students who said that they used PCs as their primary computers. Suppose we select a student at random from among those who were part of the census and learn that the student mainly uses a PC. Find the probability that this person is a graduate student. Show your work.

73. **Free downloads?** Illegal music downloading has become a big problem: 29% of Internet users download music files, and 67% of downloaders say they don't care if the music is copyrighted.[17] What percent of Internet users download music and don't care if it's copyrighted? Write the information given in terms of probabilities, and use the general multiplication rule.

pg 322

74. **At the gym** Suppose that 10% of adults belong to health clubs, and 40% of these health club members go to the club at least twice a week. What percent of all adults go to a health club at least twice a week? Write the information given in terms of probabilities, and use the general multiplication rule.

75. **Box of chocolates** According to Forrest Gump, "Life is like a box of chocolates. You never know what you're gonna get." Suppose a candy maker offers a special "Gump box" with 20 chocolate candies that look the same. In fact, 14 of the candies have soft centers and 6 have hard centers. Choose 2 of the candies from a Gump box at random.

(a) Draw a tree diagram that shows the sample space of this chance process.

(b) Find the probability that one of the chocolates has a soft center and the other one doesn't.

76. **Inspecting switches** A shipment contains 10,000 switches. Of these, 1000 are bad. An inspector draws 2 switches at random, one after the other.

(a) Draw a tree diagram that shows the sample space of this chance process.

(b) Find the probability that both switches are defective.

77. **Fill 'er up!** In a recent month, 88% of automobile drivers filled their vehicles with regular gasoline, 2% purchased midgrade gas, and 10% bought premium gas.[18] Of those who bought regular gas, 28% paid with a credit card; of customers who bought midgrade and premium gas, 34% and 42%, respectively, paid with a credit card. Suppose we select a customer at random.

pg 324

(a) Draw a tree diagram to represent this situation.

(b) Find the probability that the customer paid with a credit card. Show your work.

(c) Given that the customer paid with a credit card, find the probability that she bought premium gas. Show your work.

78. **Urban voters** The voters in a large city are 40% white, 40% black, and 20% Hispanic. (Hispanics may be of any race in official statistics, but here we are speaking of political blocks.) A mayoral candidate anticipates attracting 30% of the white vote, 90% of the black vote, and 50% of the Hispanic vote. Suppose we select a voter at random.

(a) Draw a tree diagram to represent this situation.

(b) Find the probability that this voter votes for the mayoral candidate. Show your work.

(c) Given that the chosen voter plans to vote for the candidate, find the probability that the voter is black. Show your work.

79. **Lactose intolerance** Lactose intolerance causes difficulty in digesting dairy products that contain lactose (milk sugar). It is particularly common among people of African and Asian ancestry. In the United States (ignoring other groups and people who consider themselves to belong to more than one race), 82% of the population is white, 14% is black, and 4% is Asian. Moreover, 15% of whites, 70% of blacks, and 90% of Asians are lactose intolerant.[19] Suppose we select a U.S. person at random.

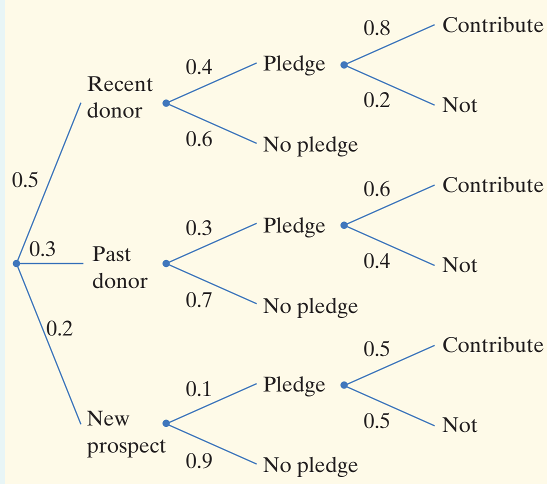
(a) What is the probability that the person is lactose intolerant? Show your work.

(b) Given that the person is lactose intolerant, find the probability that he or she is Asian. Show your work.

80. **Fundraising by telephone** Tree diagrams can organize problems having more than two stages. The figure at top right shows probabilities for a charity calling potential donors by telephone.[20] Each person called is either a recent donor, a past donor, or a new prospect. At the next stage, the person called either does or does not pledge to contribute, with conditional probabilities that depend on the donor class to which the person belongs. Finally, those who make a pledge either do or don't actually make a contribution. Suppose we randomly select a person who is called by the charity.

(a) What is the probability that the person contributed to the charity? Show your work.

(b) Given that the person contributed, find the probability that he or she is a recent donor. Show your work.



81. pg 325 **HIV testing** Enzyme immunoassay (EIA) tests are used to screen blood specimens for the presence of antibodies to HIV, the virus that causes AIDS. Antibodies indicate the presence of the virus. The test is quite accurate but is not always correct. Here are approximate probabilities of positive and negative EIA outcomes when the blood tested does and does not actually contain antibodies to HIV:[21]

	Test Result	
Truth	+	−
Antibodies present	0.9985	0.0015
Antibodies absent	0.006	0.994

Suppose that 1% of a large population carries antibodies to HIV in their blood. We choose a person from this population at random. Given that the EIA test is positive, find the probability that the person has the antibody. Show your work.

82. **Testing the test** Are false positives too common in some medical tests? Researchers conducted an experiment involving 250 patients with a medical condition and 750 other patients who did not have the medical condition. The medical technicians who were reading the test results were unaware that they were subjects in an experiment.

(a) Technicians correctly identified 240 of the 250 patients with the condition. They also identified 50 of the healthy patients as having the condition. What were the false positive and false negative rates for the test?

(b) Given that a patient got a positive test result, what is the probability that the patient actually had the medical condition? Show your work.

83. **Get rich** Refer to Exercise 63.

(a) Find P("a good chance" | female).

(b) Find P("a good chance").

(c) Use your answers to (a) and (b) to determine whether the events "a good chance" and "female" are independent. Explain your reasoning.

84. **A *Titanic* disaster** Refer to Exercise 64.

(a) Find P(survived | second class).

(b) Find P(survived).

(c) Use your answers to (a) and (b) to determine whether the events "survived" and "second class" are independent. Explain your reasoning.

pg 327 85. **Sampling senators** Refer to Exercise 65. Are events D and F independent? Justify your answer.

86. **Who eats breakfast?** Refer to Exercise 66. Are events B and M independent? Justify your answer.

87. **Rolling dice** Suppose you roll two fair, six-sided dice—one red and one green. Are the events "sum is 7" and "green die shows a 4" independent? Justify your answer.

88. **Rolling dice** Suppose you roll two fair, six-sided dice—one red and one green. Are the events "sum is 8" and "green die shows a 4" independent? Justify your answer.

89. **Bright lights?** A string of Christmas lights contains
pg 329 20 lights. The lights are wired in series, so that if any light fails, the whole string will go dark. Each light has probability 0.02 of failing during a 3-year period. The lights fail independently of each other. Find the probability that the string of lights will remain bright for 3 years.

90. **Common names** The Census Bureau says that the 10 most common names in the United States are (in order) Smith, Johnson, Williams, Brown, Jones, Miller, Davis, Garcia, Rodriguez, and Wilson. These names account for 9.6% of all U.S. residents. Out of curiosity, you look at the authors of the textbooks for your current courses. There are 9 authors in all. Would you be surprised if none of the names of these authors were among the 10 most common? (Assume that authors' names are independent and follow the same probability distribution as the names of all residents.)

91. **Universal blood donors** People with type O-nega-
pg 329 tive blood are universal donors. That is, any patient can receive a transfusion of O-negative blood. Only 7.2% of the American population have O-negative blood. If we choose 10 Americans at random who gave blood, what is the probability that at least 1 of them is a universal donor?

92. **Lost Internet sites** Internet sites often vanish or move, so that references to them can't be followed. In fact, 13% of Internet sites referenced in major scientific journals are lost within two years after publication.[22] If we randomly select seven Internet references, from scientific journals, what is the probability that at least one of them doesn't work two years later?

93. **Late shows** Some TV shows begin after their scheduled times when earlier programs run late. According to a network's records, about 3% of its shows start late. To find the probability that three consecutive shows on this network start on time, can we multiply (0.97)(0.97)(0.97)? Why or why not?

94. **Late flights** An airline reports that 85% of its flights arrive on time. To find the probability that its next four flights into LaGuardia Airport all arrive on time, can we multiply (0.85)(0.85)(0.85)(0.85)? Why or why not?

95. **The geometric distributions** You are tossing a pair of fair, six-sided dice in a board game. Tosses are independent. You land in a danger zone that requires you to roll doubles (both faces showing the same number of spots) before you are allowed to play again. How long will you wait to play again?

(a) What is the probability of rolling doubles on a single toss of the dice? (If you need review, the possible outcomes appear in Figure 5.2 (page 306). All 36 outcomes are equally likely.)

(b) What is the probability that you do not roll doubles on the first toss, but you do on the second toss?

(c) What is the probability that the first two tosses are not doubles and the third toss is doubles? This is the probability that the first doubles occurs on the third toss.

(d) Now you see the pattern. What is the probability that the first doubles occurs on the fourth toss? On the fifth toss? Give the general result: what is the probability that the first doubles occurs on the kth toss?

96. **The probability of a flush** A poker player holds a flush when all 5 cards in the hand belong to the same suit. We will find the probability of a flush when 5 cards are dealt. Remember that a deck contains 52 cards, 13 of each suit, and that when the deck is well shuffled, each card dealt is equally likely to be any of those that remain in the deck.

(a) We will concentrate on spades. What is the probability that the first card dealt is a spade? What is the conditional probability that the second card is a spade given that the first is a spade?

(b) Continue to count the remaining cards to find the conditional probabilities of a spade on the third, the fourth, and the fifth card given in each case that all previous cards are spades.

(c) The probability of being dealt 5 spades is the product of the five probabilities you have found. Why? What is this probability?

(d) The probability of being dealt 5 hearts or 5 diamonds or 5 clubs is the same as the probability of being dealt 5 spades. What is the probability of being dealt a flush?

Multiple choice:* *Select the best answer for Exercises 97 to 99.

97. An athlete suspected of using steroids is given two tests that operate independently of each other. Test A has probability 0.9 of being positive if steroids have been used. Test B has probability 0.8 of being positive if steroids have been used. What is the probability that neither test is positive if steroids have been used?

(a) 0.72 (c) 0.02 (e) 0.08

(b) 0.38 (d) 0.28

98. In an effort to find the source of an outbreak of food poisoning at a conference, a team of medical detectives carried out a study. They examined all 50 people who had food poisoning and a random sample of 200 people attending the conference who didn't get food poisoning. The detectives found that 40% of the people with food poisoning went to a cocktail party on the second night of the conference, while only 10% of the people in the random sample attended the same party. Which of the following statements is appropriate for describing the 40% of people who went to the party? (Let F = got food poisoning and A = attended party.)

(a) $P(F \mid A) = 0.40$ (d) $P(A^C \mid F) = 0.40$

(b) $P(A \mid F^C) = 0.40$ (e) $P(A \mid F) = 0.40$

(c) $P(F \mid A^C) = 0.40$

99. Suppose a loaded die has the following probability model:

Outcome:	1	2	3	4	5	6
Probability:	0.3	0.1	0.1	0.1	0.1	0.3

If this die is thrown and the top face shows an odd number, what is the probability that the die shows a 1?

(a) 0.10 (d) 0.50

(b) 0.17 (e) 0.60

(c) 0.30

Exercises 100 and 101 refer to the following setting. Your body mass index (BMI) is your weight in kilograms divided by the square of your height in meters. Online BMI calculators allow you to enter weight in pounds and height in inches. High BMI is a common but controversial indicator of being overweight or obese. A study by the National Center for Health Statistics found that the BMI of American young women (ages 20 to 29) is approximately Normal with mean 26.8 and standard deviation 7.4.[23]

100. **BMI** (2.2) People with BMI less than 18.5 are often classed as "underweight." What percent of young women are underweight by this criterion? Sketch and shade the area of interest under a Normal curve.

101. **BMI** (5.2) Suppose we select two American young women in this age group at random. Find the probability that at least one of them is classified as underweight. Show your work.

102. **Life at work** (1.1) The University of Chicago's General Social Survey asked a representative sample of adults this question: "Which of the following statements best describes how your daily work is organized? (1) I am free to decide how my daily work is organized. (2) I can decide how my daily work is organized, within certain limits. (3) I am not free to decide how my daily work is organized." Here is a two-way table of the responses for three levels of education:[24]

	Highest Degree Completed		
Response	**Less than High School**	**High School**	**Bachelor's**
1	31	161	81
2	49	269	85
3	47	112	14

Do these data suggest that there is an association between level of education and freedom to organize one's work in the adult population? Give appropriate evidence to support your answer.

FRAPPY! Free Response AP® Problem, Yay!

The following problem is modeled after actual AP® Statistics exam free response questions. Your task is to generate a complete, concise response in 15 minutes.

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

A statistics teacher has 40 students in his class, 23 females and 17 males. At the beginning of class on a Monday, the teacher planned to spend time reviewing an assignment due that day. Unknown to the teacher, only 19 of the females and 11 of the males had completed the assignment. The teacher plans to randomly select students to do problems from the assignment on the whiteboard.

(a) What is the probability that a randomly selected student has completed the assignment?

(b) Are the events "selecting a female" and "selecting a student who completed the assignment" independent? Justify your answer.

Suppose that the teacher randomly selects 4 students to do a problem on the whiteboard and only 2 of the students had completed the assignment.

(c) Describe how to use a table of random digits to estimate the probability that 2 or fewer of the 4 randomly selected students completed the assignment.

(d) Complete three repetitions of your simulation using the random digits below and use the results to estimate the probability described in part (c).

12975	13258	13048	45144	72321	81940	00360	02428
96767	35964	23822	96012	94951	65194	50842	55372
37609	59057	66967	83401	60705	02384	90597	93600

After you finish, you can view two example solutions on the book's Web site (www.whfreeman.com/tps5e). Determine whether you think each solution is "complete," "substantial," "developing," or "minimal." If the solution is not complete, what improvements would you suggest to the student who wrote it? Finally, your teacher will provide you with a scoring rubric. Score your response and note what, if anything, you would do differently to improve your own score.

Chapter Review

Section 5.1: Randomness, Probability, and Simulation

In this section, you learned about the law of large numbers and the idea of probability. The law of large numbers says that when you repeat a chance process many, many times, the relative frequency of an outcome will approach a single number. This single number is called the probability of the outcome—how often we expect the outcome to occur in a very large number of repetitions of the chance process. Make sure to remember the "large" part of the law of large numbers. Although clear patterns emerge in a large number of repetitions, we shouldn't expect such regularity in a small number of repetitions.

Simulations are powerful tools that we can use to imitate chance processes and estimate probabilities. To perform a simulation, use the familiar four-step process: state the question of interest, plan how to use a chance device to imitate a process, do many repetitions, and make a conclusion based on the results. If you are using random digits to perform your simulation, be sure to consider whether or not digits can be repeated within each trial.

Section 5.2: Probability Rules

In this section, you learned that chance behavior can be described by a probability model. Probability models have two parts, a list of possible outcomes (the sample space) and a probability for each outcome. The probability of each outcome in a probability model must be between 0 and 1, and the probabilities of all the outcomes in the sample space must add to 1.

An event is a subset of the possible outcomes of a chance process. The complement rule says that the probability that an event occurs is 1 minus the probability that the event doesn't occur. In symbols, the complement rule says that $P(E) = 1 - P(E^C)$. Given two events A and B from some chance process, use the general addition rule to find the probability that event A or event B occurs:

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If the events A and B have no outcomes in common, use the addition rule for mutually exclusive events: $P(A \cup B) = P(A) + P(B)$.

Finally, you learned how to use two-way tables and Venn diagrams to display the sample space for a chance process involving two events. Using a two-way table or a Venn diagram is a helpful way to organize information and calculate probabilities involving the union (A or B) and the intersection (A and B) of two events.

Section 5.3: Conditional Probability and Independence

In this section, you learned that a conditional probability describes the probability of an event occurring given that another event is known to have already occurred. To calculate the probability that event A occurs given that event B has occurred, use the formula

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \text{ and } B)}{P(B)}$$

Two-way tables and tree diagrams are useful ways to organize the information provided in a conditional probability problem. Two-way tables are best when the problem describes the number or proportion of cases with certain characteristics. Tree diagrams are best when the problem provides the conditional probabilities of different events or describes a sequence of events.

Use the general multiplication rule for calculating the probability that event A and event B both occur:

$$P(A \text{ and } B) = P(A \cap B) = P(A) \cdot P(B \mid A)$$

If knowing that event B occurs doesn't change the probability that event A occurs, then events A and B are independent. That is, events A and B are independent if $P(A \mid B) = P(A)$. If events A and B are independent, use the multiplication rule for independent events to find the probability that events A and B both occur: $P(A \cap B) = P(A) \cdot P(B)$.

What Did You Learn?

Learning Objective	Section	Related Example on Page(s)	Relevant Chapter Review Exercise(s)
Interpret probability as a long-run relative frequency.	5.1	291, 292, 293, 294, 295	R5.1
Use simulation to model chance behavior.	5.1	296, 297	R5.2
Determine a probability model for a chance process.	5.2	306	R5.3, R5.10
Use basic probability rules, including the complement rule and the addition rule for mutually exclusive events.	5.2	308	R5.4, R5.10
Use a two-way table or Venn diagram to model a chance process and calculate probabilities involving two events.	5.2	309, 312, 313	R5.4, R5.5, R5.7, R5.8
Use the general addition rule to calculate probabilities.	5.2	313	R5.4, R5.5
Calculate and interpret conditional probabilities.	5.3	318, 320	R5.6, R5.8
Use the general multiplication rule to calculate probabilities.	5.3	322	R5.6
Use tree diagrams to model a chance process and calculate probabilities involving two or more events.	5.3	322, 324, 325	R5.6
Determine whether two events are independent.	5.3	327	R5.7, R5.8
When appropriate, use the multiplication rule for independent events to compute probabilities.	5.3	329, 331	R5.9, R5.10

Chapter 5 Chapter Review Exercises

These exercises are designed to help you review the important ideas and methods of the chapter.

R5.1 Rainy days The TV weatherman says, "There's a 30% chance of rain tomorrow." Explain what this statement means.

R5.2 Click it or else From police records, it has been determined that 15% of drivers stopped for routine license checks are not wearing seat belts. If a police officer stops 10 vehicles, how likely is it that two consecutive drivers won't be wearing their seat belts?

(a) Describe the design of a simulation to estimate this probability. Explain clearly how you will use the partial table of random digits below to carry out your simulation.

(b) Carry out three repetitions of the simulation. Copy the random digits below onto your paper. Then mark on or directly above the table to show your results.

29077	14863	61683	47052	62224	51025
95052	90908	73592	75186	87136	95761
27102	56027	55892	33063	41842	81868
43367	49497	72719	96758	27611	91596

R5.3 Weird dice Nonstandard dice can produce interesting distributions of outcomes. Suppose you have two balanced, six-sided dice. Die A has faces with 2, 2, 2, 2, 6, and 6 spots. Die B has faces with 1, 1, 1, 5, 5, and 5 spots. Imagine that you roll both dice at the same time.

(a) Find a probability model for the difference (Die A − Die B) in the total number of spots on the up-faces.

(b) Which die is more likely to roll a higher number? Justify your answer.

R5.4 Race and ethnicity The Census Bureau allows each person to choose from a long list of races. That is, in the eyes of the Census Bureau, you belong to whatever race you say you belong to. Hispanic (also called Latino) is a separate category. Hispanics may be of any race. If we choose a resident of the United States at random, the Census Bureau gives these probabilities:[25]

	Hispanic	Not Hispanic
Asian	0.001	0.044
Black	0.006	0.124
White	0.139	0.674
Other	0.003	0.009

(a) Verify that this is a legitimate assignment of probabilities.

(b) What is the probability that a randomly chosen American is Hispanic?

(c) Non-Hispanic whites are the historical majority in the United States. What is the probability that a randomly chosen American is *not* a member of this group?

(d) Explain why $P(\text{white or Hispanic}) \neq P(\text{white}) + P(\text{Hispanic})$. Then find $P(\text{white or Hispanic})$.

R5.5 In 2012, fans at Arizona Diamondbacks home games would win 3 free tacos from Taco Bell if the Diamondbacks scored 6 or more runs. In the 2012 season, the Diamondbacks won 41 of their 81 home games and gave away free tacos in 30 of their 81 home games. In 26 of the games, the Diamondbacks won and gave away free tacos. Choose a Diamondbacks home game at random.

(a) Make a Venn diagram to model this chance process.

(b) What is the probability that the Diamondbacks lost and did not give away free tacos?

(c) What is the probability that the Diamondbacks won the game or fans got free tacos?

R5.6 Steroids A company has developed a drug test to detect steroid use by athletes. The test is accurate 95% of the time when an athlete has taken steroids. It is 97% accurate when an athlete hasn't taken steroids. Suppose that the drug test will be used in a population of athletes in which 10% have actually

taken steroids. Let's choose an athlete at random and administer the drug test.

(a) Make a tree diagram showing the sample space of this chance process.
(b) What's the probability that the randomly selected athlete tests positive? Show your work.
(c) Suppose that the chosen athlete tests positive. What's the probability that he or she actually used steroids? Show your work.

R5.7 **Mike's pizza** You work at Mike's pizza shop. You have the following information about the 7 pizzas in the oven: 3 of the 7 have thick crust and 2 of the 3 thick crust pizzas have mushrooms. Of the remaining 4 pizzas, 2 have mushrooms. Choose a pizza at random from the oven.

(a) Make a two-way table to model this chance process.
(b) Are the events "getting a thick-crust pizza" and "getting a pizza with mushrooms" independent? Explain.
(c) You add an eighth pizza to the oven. This pizza has thick crust with only cheese. Now are the events "getting a thick-crust pizza" and "getting a pizza with mushrooms" independent? Explain.

R5.8 **Deer and pine seedlings** As suburban gardeners know, deer will eat almost anything green. In a study of pine seedlings at an environmental center in Ohio, researchers noted how deer damage varied with how much of the seedling was covered by thorny undergrowth:[26]

	Deer Damage	
Thorny Cover	Yes	No
None	60	151
$<1/3$	76	158
1/3 to 2/3	44	177
$>2/3$	29	176

(a) What is the probability that a randomly selected seedling was damaged by deer?

(b) What are the conditional probabilities that a randomly selected seedling was damaged, given each level of cover?

(c) Does knowing about the amount of thorny cover on a seedling change the probability of deer damage? Justify your answer.

R5.9 **A random walk on Wall Street?** The "random walk" theory of stock prices holds that price movements in disjoint time periods are independent of each other. Suppose that we record only whether the price is up or down each year, and that the probability that our portfolio rises in price in any one year is 0.65. (This probability is approximately correct for a portfolio containing equal dollar amounts of all common stocks listed on the New York Stock Exchange.)

(a) What is the probability that our portfolio goes up for three consecutive years?

(b) What is the probability that the portfolio's value moves in the same direction (either up or down) for three consecutive years?

R5.10 **Blood types** Each of us has an ABO blood type, which describes whether two characteristics called A and B are present. Every human being has two blood type alleles (gene forms), one inherited from our mother and one from our father. Each of these alleles can be A, B, or O. The two that we inherit determine our blood type. The table shows what our blood type is for each combination of two alleles. We inherit each of a parent's two alleles with probability 0.5. We inherit independently from our mother and father.

Alleles inherited	Blood type
A and A	A
A and B	AB
A and O	A
B and B	B
B and O	B
O and O	O

(a) Hannah and Jacob both have alleles A and B. Diagram the sample space that shows the alleles that their next child could receive. Then give the possible blood types that this child could have, along with the probability for each blood type.

(b) Jennifer has alleles A and O. Jose has alleles A and B. They have two children. What is the probability that at least one of the two children has blood type B? Show your method.

Chapter 5 AP® Statistics Practice Test

Section I: Multiple Choice *Select the best answer for each question.*

T5.1 Dr. Stats plans to toss a fair coin 10,000 times in the hope that it will lead him to a deeper understanding of the laws of probability. Which of the following statements is true?

(a) It is unlikely that Dr. Stats will get more than 5000 heads.
(b) Whenever Dr. Stats gets a string of 15 tails in a row, it becomes more likely that the next toss will be a head.
(c) The fraction of tosses resulting in heads should be exactly 1/2.
(d) The chance that the 100th toss will be a head depends somewhat on the results of the first 99 tosses.
(e) It is likely that Dr. Stats will get about 50% heads.

T5.2 China has 1.2 billion people. Marketers want to know which international brands they have heard of. A large study showed that 62% of all Chinese adults have heard of Coca-Cola. You want to simulate choosing a Chinese at random and asking if he or she has heard of Coca-Cola. One correct way to assign random digits to simulate the answer is:

(a) One digit simulates one person's answer; odd means "Yes" and even means "No."
(b) One digit simulates one person's answer; 0 to 6 mean "Yes" and 7 to 9 mean "No."
(c) One digit simulates the result; 0 to 9 tells how many in the sample said "Yes."
(d) Two digits simulate one person's answer; 00 to 61 mean "Yes" and 62 to 99 mean "No."
(e) Two digits simulate one person's answer; 00 to 62 mean "Yes" and 63 to 99 mean "No."

T5.3 Choose an American household at random and record the number of vehicles they own. Here is the probability model if we ignore the few households that own more than 5 cars:

Number of cars:	0	1	2	3	4	5
Probability:	0.09	0.36	0.35	0.13	0.05	0.02

A housing company builds houses with two-car garages. What percent of households have more cars than the garage can hold?

(a) 7% (b) 13% (c) 20% (d) 45% (e) 55%

T5.4 Computer voice recognition software is getting better. Some companies claim that their software correctly recognizes 98% of all words spoken by a trained user. To simulate recognizing a single word when the probability of being correct is 0.98, let two digits simulate one word; 00 to 97 mean "correct." The program recognizes words (or not) independently. To simulate the program's performance on 10 words, use these random digits:

60970 70024 17868 29843 61790 90656 87964

The number of words recognized correctly out of the 10 is

(a) 10 (b) 9 (c) 8 (d) 7 (e) 6

Questions T5.5 to T5.7 refer to the following setting. One thousand students at a city high school were classified according to both GPA and whether or not they consistently skipped classes. The two-way table below summarizes the data. Suppose that we choose a student from the school at random.

	GPA		
Skipped Classes	**<2.0**	**2.0–3.0**	**>3.0**
Many	80	25	5
Few	175	450	265

T5.5 What is the probability that a student has a GPA under 2.0?

(a) 0.227 (b) 0.255 (c) 0.450 (d) 0.475 (e) 0.506

T5.6 What is the probability that a student has a GPA under 2.0 or has skipped many classes?

(a) 0.080 (b) 0.281 (c) 0.285 (d) 0.365 (e) 0.727

T5.7 What is the probability that a student has a GPA under 2.0 given that he or she has skipped many classes?

(a) 0.080 (b) 0.281 (c) 0.285 (d) 0.314 (e) 0.727

T5.8 For events A and B related to the same chance process, which of the following statements is true?

(a) If A and B are mutually exclusive, then they must be independent.
(b) If A and B are independent, then they must be mutually exclusive.
(c) If A and B are not mutually exclusive, then they must be independent.
(d) If A and B are not independent, then they must be mutually exclusive.
(e) If A and B are independent, then they cannot be mutually exclusive.

T5.9 Choose an American adult at random. The probability that you choose a woman is 0.52. The probability that the person you choose has never married is 0.25. The probability that you choose a woman who has never married is 0.11. The probability that the person you choose is either a woman or has never been married (or both) is therefore about

(a) 0.77. **(b)** 0.66. **(c)** 0.44. **(d)** 0.38. **(e)** 0.13.

T5.10 A deck of playing cards has 52 cards, of which 12 are face cards. If you shuffle the deck well and turn over the top 3 cards, one after the other, what's the probability that all 3 are face cards?

(a) 0.001 **(c)** 0.010 **(e)** 0.02
(b) 0.005 **(d)** 0.012

Section II: Free Response *Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.*

T5.11 Your teacher has invented a "fair" dice game to play. Here's how it works. Your teacher will roll one fair *eight-sided* die, and you will roll a fair *six-sided* die. Each player rolls once, and the winner is the person with the higher number. In case of a tie, neither player wins. The table shows the sample space of this chance process.

	Teacher Rolls							
You Roll	1	2	3	4	5	6	7	8
1								
2								
3								
4								
5								
6								

(a) Let A be the event "your teacher wins." Find $P(A)$.
(b) Let B be the event "you get a 3 on your first roll." Find $P(A \cup B)$.
(c) Are events A and B independent? Justify your answer.

T5.12 Three machines—A, B, and C—are used to produce a large quantity of identical parts at a factory. Machine A produces 60% of the parts, while Machines B and C produce 30% and 10% of the parts, respectively. Historical records indicate that 10% of the parts produced by Machine A are defective, compared with 30% for Machine B and 40% for Machine C.

(a) Draw a tree diagram to represent this chance process.
(b) If we choose a part produced by one of these three machines, what's the probability that it's defective? Show your work.
(c) If a part is inspected and found to be defective, which machine is most likely to have produced it? Give appropriate evidence to support your answer.

T5.13 Researchers are interested in the relationship between cigarette smoking and lung cancer. Suppose an adult male is randomly selected from a particular population. The following table shows the probabilities of some events related to this chance process:

Event	Probability
Smokes	0.25
Smokes and gets cancer	0.08
Does not smoke and does not get cancer	0.71

(a) Find the probability that the individual gets cancer given that he is a smoker. Show your work.
(b) Find the probability that the individual smokes or gets cancer. Show your work.
(c) Two adult males are selected at random. Find the probability that at least one of the two gets cancer. Show your work.

T5.14 Based on previous records, 17% of the vehicles passing through a tollbooth have out-of-state plates. A bored tollbooth worker decides to pass the time by counting how many vehicles pass through until he sees two with out-of-state plates.[27]

(a) Describe the design of a simulation to estimate the average number of vehicles it takes to find two with out-of-state plates. Explain clearly how you will use the partial table of random digits below to carry out your simulation.
(b) Perform three repetitions of the simulation you described in part (a). Copy the random digits below onto your paper. Then mark on or directly above the table to show your results.

41050	92031	06449	05059	59884	31880
53115	84469	94868	57967	05811	84514
84177	06757	17613	15582	51506	81435
75011	13006	63395	55041	15866	06589

AP1.12 b
AP1.13 b
AP1.14 a
AP1.15 **(a)** The distribution of gains for subjects using Machine A is roughly symmetric, while the distribution of gains for subjects using Machine B is skewed to the left. The center is greater for Machine B than for Machine A. The distribution for Machine B is more variable than the distribution for Machine A. **(b)** B. The typical gain using Machine B is greater than the typical gain using Machine A. **(c)** A. The spread for Machine A is less than the spread for Machine B. **(d)** Volunteers from one fitness center were used and these volunteers may be different in some way from the general population of those who are interested in cardiovascular fitness. To broaden their scope of inference, they should randomly select people from the population they would like to draw an inference about.
AP1.16 **(a)** Number the 60 retail sales districts with a two-digit number from 01 to 60. Using a table of random digits, read two-digit numbers until 30 unique numbers from 01 to 60 have been selected. The corresponding 30 districts are assigned to the monetary incentives group and the remaining 30 to the tangible incentives group. After a specified period of time, record the change in sales for each district and compare the two groups. **(b)** The districts labeled 07, 51, and 18 are the first three to be assigned to the monetary incentives group. **(c)** Pair the two districts with the largest sales, the next two largest, down to the two smallest districts. For each pair, pick one of the districts and flip a coin. If the flip is "heads," this district is assigned to the monetary incentives group. If it is "tails," this district is assigned to the tangible incentives group. The other district in the pair is assigned to the other group. After a specified period of time, record the change in sales for each district and compare within each pair.
AP1.17 **(a)** There is a very strong, positive, linear association between sales and shelf length. **(b)** $\hat{y} = 317.94 + 152.68x$, where y = weekly sales (in dollars) and x = shelf length (in feet). **(c)** \$1081 **(d)** When using the least-squares regression line with x = shelf space to predict y = sales, we will typically be off by about s = \$23. **(e)** \$\$\$ About 98.2% of the variation in weekly sales revenue can be accounted for by the linear model relating sales to shelf length. **(f)** It would be inappropriate to interpret the intercept, because the data represent sales based on shelf lengths of 3 to 6 feet and 0 feet falls substantially outside that domain.

Chapter 5

Section 5.1

Answers to Check Your Understanding

page 292: **1.** **(a)** If you asked a large sample of U.S. adults whether they usually eat breakfast, about 61% of them will answer yes. **(b)** The exact number of breakfast eaters will vary from sample to sample. **2.** **(a)** 0. If an outcome can never occur, then it will occur in 0% of the trials. **(b)** 1. If an outcome will occur on every trial, then it will occur in 100% of the trials. **(c)** 0.01. An outcome that occurs in 1% of the trials is very unlikely, but will occur every once in a while. **(d)** 0.6. An outcome that occurs in 60% of the trials will happen more than half of the time.

page 299: **1.** Assign the members of the AP® Statistics class the numbers 01–28 and the rest of the students numbers 29–95. Ignore the numbers 96–99 and 00. In Table D, read off 4 two-digit numbers, making sure that the second number is different than the first and that the fourth number is different than the third. Record whether all four numbers are between 01 and 28 or not. **2.** Assign the numbers 1–10 to Jeff Gordon, 11–40 to Dale Earnhardt, Jr., 41–60 to Tony Stewart, 61–85 to Danica Patrick, and 86–100 to Jimmie Johnson. Then proceed as in the example.

Answers to Odd-Numbered Section 5.1 Exercises

5.1 **(a)** If we use a polygraph machine on many, many people who are all telling the truth, the machine will say about 8% of the people are lying. **(b)** Answers will vary. A false positive would mean that a person telling the truth would be found to be lying. A false negative would mean that a person lying would be found to be telling the truth.
5.3 **(a)** If we look at many families like this, approximately 25% of them will have a first-born child that develops cystic fibrosis. **(b)** No. The number of children with cystic fibrosis could be smaller or larger than 4 by random chance.
5.5 **(a)** Answers will vary. **(b)** Spin the coin many more times.
5.7 In the short run, there was quite a bit of variability in the percentage of made free throws. However, this percentage became less variable and approached 0.30 as the number of shots increased.
5.9 No, he is incorrectly applying the law of large numbers to a small number of at-bats.
5.11 **(a)** There are 10,000 four-digit numbers (0000, 0001, . . . , 2873, . . . , 9999), and each is equally likely. **(b)** 2873. To many, 2873 "looks" more random than 9999—we don't "expect" to get the same number four times in a row. It would be best to choose a number that others would avoid so you don't have to split the pot with many other people.
5.13 **(a)** Let diamonds, spades, and clubs represent making a free throw and hearts represent missing. Deal one card from the deck. **(b)** Let 00–74 represent making the free throw and 75–99 represent missing. Read a two-digit number from Table D. **(c)** Let 1–3 represent the player making the free throw and 4 represent a miss. Generate a random integer from 1–4.
5.15 **(a)** There are 19 (not 18) numbers from 00 to 18, 19 (not 18) numbers from 19 to 37, and 3 (not 2) numbers from 38 to 40. **(b)** Repeats should not be skipped. For example, if the first number selected was 08, then the probability of selecting a left-hander on the next selection would be 9% (instead of 10%).
5.17 **(a)** Valid. The chance of rolling a 1, 2, or 3 is 75% on a 4-sided die and the 100 rolls represent the 100 randomly selected U.S. adults. **(b)** Not valid. The probability of heads is 50% rather than 60%. This method will underestimate the number of times she hits the center of the target.
5.19 **(a)** What is the probability that, in a random selection of 10 passengers, none from first class are chosen? **(b)** Number the first-class passengers 01–12 and the other passengers 13–76. Look up two-digit numbers in Table D until you have 10 unique numbers from 01 to 76. Count the numbers between 01 and 12. **(c)** 71 48 70 ~~99 84~~ 29 07 ~~71 48~~ 63 61 68 34 ~~70~~ 52. There is one person selected who is in first class. **(d)** It seems plausible that the actual selection was random, because 15/100 is not very small.
5.21 **(a)** Use a random integer generator to select 30 numbers from 1 to 365. Record whether or not there were any repeats in the sample. **(b)** Answers will vary. **(c)** Answers will vary.
5.23 **(a)** Obtaining a sample percentage of 55% or higher is not particularly unusual (probability ≈ 43/200) when 50% of all students recycle. **(b)** Obtaining a sample percentage of at least 63% is very unlikely (probability ≈ 1/200) when 50% of all students recycle.

5.25 *State*: What is the probability that, in a sample of 4 randomly selected U.S. adult males, at least one of them is red-green colorblind? *Plan*: Let 00–06 denote a colorblind man and 07–99 denote a non-colorblind man. Read 4 two-digit numbers from Table D for each sample and record whether or not the sample had at least one red-green colorblind man in it. *Do*: In our 50 samples, 15 had at least one colorblind man in them. *Conclude*: The probability that a sample of 4 men would have at least one colorblind man is approximately 15/50 = 0.30.
5.27 *State*: What is the probability that it takes 20 or more selections in order to find one man who is red-green colorblind? *Plan*: Let 0–6 denote a colorblind man and 7–99 denote a non-colorblind man. Use technology to pick integers from 0 to 99 until we get a number between 0 and 6. Count how many numbers there are in the sample. *Do*: In 16 of our 50 samples, it took 20 or more selections to get one colorblind man. *Conclude*: Not surprised. The probability of needing 20 or more selections to get one colorblind man is fairly large (approximately 16/50 = 0.32).
5.29 *State*: What is the probability that the random assignment will result in at least 6 men in the same group? *Plan*: Number the men 1–8 and the women 9–20. Use technology to pick 10 unique integers between 1 and 20 for one group. Record if there are at least 6 numbers between 1 and 8 in either group. *Do*: In our 50 repetitions, 9 had one group with 6 or more men in it. *Conclude*: Not surprised. The probability of getting 6 or more men in one group is fairly large (approximately 9/50 = 0.18).
5.31 c
5.33 b
5.35 c
5.37 **(a)** Population: adult U.S. residents. Sample: the 353,564 adults who were interviewed. **(b)** The people who do not have a telephone were excluded. This would lead to an underestimate of the proportion in the population who experienced stress a lot of the day yesterday if the people without phones are poorer and consequently experience more stress.

Section 5.2

Answers to Check Your Understanding

page 309: **1.** A person cannot have a cholesterol level of both 240 or above and between 200 and 239 at the same time. **2.** A person has either a cholesterol level of 240 or above, or they have a cholesterol level between 200 and 239. $P(A \text{ or } B) = 0.16 + 0.29 = 0.45$. **3.** $P(C) = 1 - 0.45 = 0.55$.
page 311: **1.**

	Face card	Non-face card	Total
Heart	3	10	**13**
Non-heart	9	30	**39**
Total	**12**	**40**	**52**

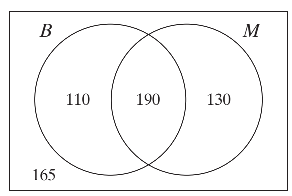
2. $P(F \text{ and } H) = 3/52 = 0.058$. **3.** The face cards that are hearts will be double-counted because F and H are not mutually exclusive. $P(F \text{ or } H) = \frac{12}{52} + \frac{13}{52} - \frac{3}{52} = \frac{22}{52} = 0.423$.

Answers to Odd-Numbered Section 5.2 Exercises

5.39 **(a)** (1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4), (4,1), (4,2), (4,3), (4,4). **(b)** Each outcome has probability $\frac{1}{16}$.
5.41 $P(A) = \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = \frac{4}{16} = 0.25$.
5.43 **(a)** Legitimate. **(b)** Not legitimate: the total is more than 1. **(c)** Legitimate.
5.45 **(a)** $P(\text{type AB}) = 1 - 0.96 = 0.04$ **(b)** $P(\text{not type AB}) = 1 - P(\text{type AB}) = 1 - 0.04 = 0.96$ **(c)** $P(\text{type O or B}) = 0.49 + 0.20 = 0.69$
5.47 **(a)** $1 - 0.13 - 0.29 - 0.30 = 0.28$ **(b)** Using the complement rule, $1 - 0.13 = 0.87$.
5.49 **(a)** $P(\text{Female}) = \frac{275}{595} = 0.462$ **(b)** $P(\text{Eats breakfast regularly}) = \frac{300}{595} = 0.504$. **(c)** $P(\text{Female and breakfast}) = \frac{110}{595} = 0.185$. **(d)** $P(\text{Female or breakfast}) = \frac{275}{595} + \frac{300}{595} - \frac{110}{595} = \frac{465}{595} = 0.782$.
5.51 **(a)**

	B	Not B	Total
E	10	10	**20**
Not E	8	10	**18**
Total	**18**	**20**	**38**

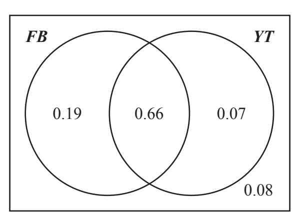
(b) $P(B) = \frac{18}{38} = 0.474$; $P(E) = \frac{20}{38} = 0.526$. **(c)** The ball lands in a spot that is black and even. $P(B \text{ and } E) = \frac{10}{38} = 0.263$. **(d)** If we add the probabilities of B and E, the spots that are black and even will be double-counted. $P(B \text{ or } E) = \frac{18}{38} + \frac{20}{38} - \frac{10}{38} = \frac{28}{38} = 0.737$.
5.53 **(a)**



(b) $P(B \cup M) = \frac{430}{595} = 0.723$. There is a 0.723 probability that we select a person who is a breakfast eater, a male, or both. **(c)** $P(B^C \cap M^C) = \frac{165}{595} = 0.277$. There is a 0.277 probability that we select a female who is not a breakfast eater.
5.55 **(a)**

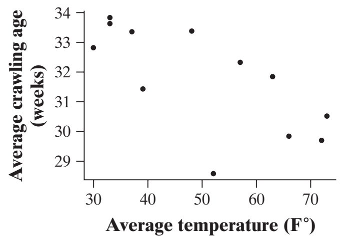
	FB	Not FB	Total
YT	0.66	0.07	**0.73**
Not YT	0.19	0.08	**0.27**
Total	**0.85**	**0.15**	**1**

(b)



(c) $FB \cup YT$ **(d)** $P(FB \cup YT) = 0.85 + 0.73 - 0.66 = 0.92$.

5.57 c
5.59 c
5.61 The scatterplot for the average crawling age and average temperature is given below.



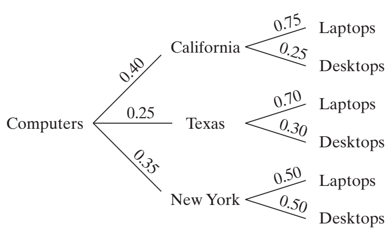
In this scatterplot, there appears to be a moderately strong, negative linear relationship between average temperature and average crawling age. The equation for the least-squares regression line is $\widehat{age} = 35.7 - 0.077\,(\text{temp})$. We predict that babies will walk 0.077 weeks earlier for every degree warmer it gets.

Section 5.3

Answers to Check Your Understanding

page 321: **1.** $P(L) = \dfrac{3656}{10{,}000} = 0.3656$. There is a 0.3656 probability of selecting a course grade that is lower than a B. **2.** $P(E \mid L) = \dfrac{800}{3656} = 0.219$. $P(L \mid E) = \dfrac{800}{1600} = 0.50$. $P(L \mid E)$ gives the probability of getting a lower grade given that the student is studying engineering or physical science. Because this probability (0.50) is greater than $P(L) = 0.3656$, we can conclude that grades are lower in engineering and physical sciences.

page 326: **1.**



2. $P(\text{laptop}) = 0.30 + 0.175 + 0.175 = 0.65$.

3. $P(\text{made in CA} \mid \text{laptop}) = \dfrac{0.30}{0.65} = 0.462$.

page 328: **1.** Independent. Because we are replacing the cards, knowing what the first card was will not help us predict what the second card will be. **2.** Not independent. Once we know the suit of the first card, then the probability of getting a heart on the second card will change depending on what the first card was. **3.** Independent. $P(\text{right-handed}) = 24/28 = 6/7$ is the same as $P(\text{right-handed} \mid \text{female}) = 18/21 = 6/7$.

page 331: **1.** $P(\text{returned safely}) = 0.95$. So $P(\text{safe return on all 20 missions}) = 0.95^{20} = 0.3585$. **2.** No. Being a college student and being 55 or older are not independent events.

Answers to Odd-Numbered Section 5.3 Exercises

5.63 **(a)** $P(\text{almost certain}|\text{M}) = \dfrac{597}{2459} = 0.2428$.

(b) $P(\text{F} \mid \text{Some chance}) = \dfrac{426}{712} = 0.5983$.

5.65 **(a)** $P(D \mid F) = \dfrac{13}{17} = 0.7647$. Given that a senator is female, there is a 0.7647 probability that she is a Democrat.

(b) $P(F \mid D) = \dfrac{13}{60} = 0.2167$. Given that a senator is a Democrat, there is a 0.2167 probability that she is a female.

5.67 **(a)** $P(\text{not English}) = 1 - 0.59 = 0.41$.

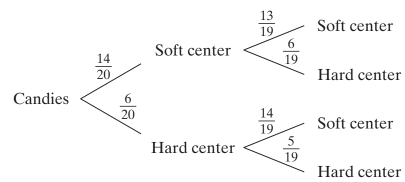
(b) $P(\text{Spanish} \mid \text{other than English}) = \dfrac{0.26}{0.41} = 0.6341$.

5.69 $P(B) < P(B \mid T) < P(T) < P(T \mid B)$. There are very few pro basketball players, so $P(B)$ should be smallest. If you are a pro basketball player, it is quite likely that you are tall, so $P(T \mid B)$ should be largest. Finally, it's much more likely to be over 6 feet tall than it is to be a pro basketball player if you're over 6 feet tall.

5.71 $P(\text{YT} \mid \text{FB}) = \dfrac{0.66}{0.85} = 0.7765$.

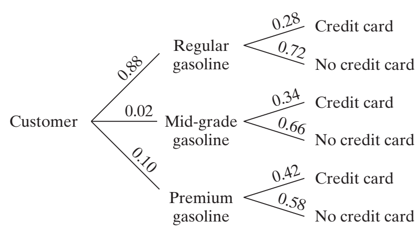
5.73 $P(\text{download music}) = 0.29$, $P(\text{don't care} \mid \text{download music}) = 0.67$.
$P(\text{download music} \cap \text{don't care}) = (0.29)(0.67) = 0.1943 = 19.43\%$.

5.75 **(a)** A tree diagram is below.



(b) $P(\text{one soft} \cap \text{one hard}) = \left(\frac{14}{20}\right)\left(\frac{6}{19}\right) + \left(\frac{6}{20}\right)\left(\frac{14}{19}\right) = \frac{168}{380} = 0.4421$

5.77 **(a)** A tree diagram is below.



(b) $P(\text{credit card}) = (0.88)(0.28) + (0.02)(0.34) + (0.10)(0.42) = 0.2952$ **(c)** $P(\text{premium gasoline}|\text{credit card}) = \dfrac{0.0420}{0.2952} = 0.142$.

5.79 **(a)** $P(\text{lactose intolerant}) = (0.82)(0.15) + (0.14)(0.70) + (0.04)(0.90) = 0.257$.

(b) $P(\text{Asian} \mid \text{lactose intolerant}) = \dfrac{0.036}{0.257} = 0.1401$.

5.81 $P(\text{antibody} \mid \text{positive}) =$

$$\frac{(0.01)(0.9985)}{(0.01)(0.9985) + (0.99)(0.006)} = 0.6270$$

5.83 **(a)** $\dfrac{663}{2367} = 0.2801$ **(b)** $\dfrac{1421}{4826} = 0.2944$ **(c)** The events are not independent because the probabilities in parts (a) and (b) are not the same.

5.85 Not independent. From Exercise 5.65, we saw that $P(D \mid F) = 0.7647$, which is not the same as $P(D) = \dfrac{60}{100} = 0.60$.

5.87 Independent. P(sum of 7 | green is 4) = 1/6 = 0.1667, which equals P(sum of 7) = 6/36 = 0.1667.
5.89 P(all remain bright) = $(0.98)^{20} = 0.6676$
5.91 P(at least one universal donor) = $1 - (0.928)^{10} = 0.5263$
5.93 No, because the events are not independent. If one show starts late, we can predict that the next show will start late as well.
5.95 **(a)** $P(\text{doubles}) = \frac{6}{36} = \frac{1}{6} = 0.167$

(b) $P(\text{no doubles first} \cap \text{doubles second}) = \frac{5}{6}\left(\frac{1}{6}\right) = \frac{5}{36} = 0.139$

(c) $P(\text{first doubles on third roll}) = \frac{5}{6}\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) = \frac{25}{216} = 0.116$

(d) 4th: $\left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right)$. 5th: $\left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right)$. The probability that the first doubles are rolled on the kth roll is $\left(\frac{5}{6}\right)^{k-1}\left(\frac{1}{6}\right)$.

5.97 c
5.99 e
5.101 P(at least one is underweight) = $1 - (1 - 0.131)^2 = 0.2448$

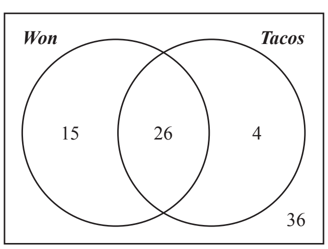
Answers to Chapter 5 Review Exercises

R5.1 When the weather conditions are like those seen today, it has rained on the following day about 30% of the time.
R5.2 **(a)** Let the numbers 00–14 represent not wearing a seat belt and 15–99 represent wearing a seat belt. Read 10 sets of two-digit numbers. For each set of 10 two-digit numbers, record whether there are two consecutive numbers between 00–14 or not. **(b)** The first sample is 29 **07** 71 48 63 61 68 34 70 52 (not two consecutive). The second sample is 62 22 45 **10** 25 95 **05** 29 **09 08** (two consecutive). The third sample is 73 59 27 51 86 87 **13** 69 57 61 (not two consecutive).
R5.3 **(a)**

Difference	1	5	–3
Probability	$\frac{18}{36}$	$\frac{6}{36}$	$\frac{12}{36}$

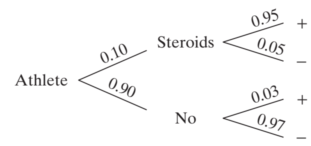
(b) Die A. $P(A > B) = P(\text{positive difference}) = \frac{18}{36} + \frac{6}{36} = \frac{24}{36}$

R5.4 **(a)** Legitimate. All probabilities are between 0 and 1 and they add up to 1. **(b)** P(Hispanic) = 0.001 + 0.006 + 0.139 + 0.003 = 0.149 **(c)** P(not a non-Hispanic white) = 1 − 0.674 = 0.326 **(d)** People who are white and Hispanic will be double-counted. P(white or Hispanic) = 0.813 + 0.149 − 0.139 = 0.823.
R5.5 **(a)**



(b) P(lost and no tacos) = 36/81 = 0.444 **(c)** $P(\text{won or tacos}) = \frac{41}{81} + \frac{30}{81} - \frac{26}{81} = \frac{45}{81} = 0.556.$

R5.6 **(a)**



(b) $P(+) = (0.10)(0.95) + (0.90)(0.03) = 0.122.$

(c) $P(\text{steroids} \mid +) = \frac{0.095}{0.122} = 0.7787.$

R5.7 **(a)**

	Thick	Thin	Total
Mushrooms	2	2	**4**
No mushrooms	1	2	**3**
Total	**3**	**4**	**7**

(b) Not independent: $P(\text{mushrooms}) = \frac{4}{7} = 0.571$ does not equal $P(\text{mushrooms} \mid \text{thick crust}) = \frac{2}{3} = 0.667.$

(c) Independent: $P(\text{mushrooms}) = \frac{4}{8} = \frac{1}{2} = 0.50$ is equal to $P(\text{mushrooms} \mid \text{thick crust}) = \frac{2}{4} = \frac{1}{2} = 0.50.$

R5.8 **(a)** $P(\text{damage}) = \frac{209}{871} = 0.24.$

(b) $P(\text{damage} \mid \text{no cover}) = \frac{60}{211} = 0.2844,$

$P\left(\text{damage} \mid < \frac{1}{3}\right) = \frac{76}{234} = 0.3248,$

$P\left(\text{damage} \mid \frac{1}{3} \text{ to } \frac{2}{3}\right) = \frac{44}{221} = 0.1991,$ and

$P\left(\text{damage} \mid > \frac{2}{3}\right) = \frac{29}{205} = 0.1415.$ **(c)** Yes. It appears that deer do much more damage when there is no cover or less than 1/3 cover than when there is more cover.

R5.9 **(a)** P(up three consecutive years) = $(0.65)^3 = 0.274625.$
(b) P(same direction for 3 years) = $(0.65)^3 + (0.35)^3 = 0.3175.$
R5.10 **(a)** (A, A), (A, B), (B, A), (B, B)

Blood type	A	AB	B
Probability	0.25	0.5	0.25

(b) (A, A), (A, B), (O, A), (O, B)

Blood type	A	AB	B
Probability	0.5	0.25	0.25

P(at least 1 type B) = $1 - P$(neither are type B) = $1 - (0.75)^2$ = 0.4375.

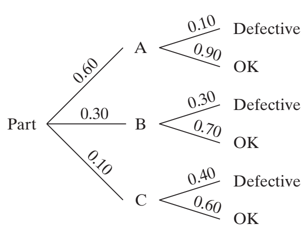
Answers to Chapter 5 AP® Statistics Practice Test

T5.1 e
T5.2 d
T5.3 c
T5.4 b
T5.5 b
T5.6 c
T5.7 e
T5.8 e
T5.9 b
T5.10 c
T5.11 **(a)** Here is a completed table, with T indicating that the teacher wins and Y indicating that you win. $P(\text{teacher wins}) = \frac{27}{48} = 0.5625$.

	1	2	3	4	5	6	7	8
1	—	T	T	T	T	T	T	T
2	Y	—	T	T	T	T	T	T
3	Y	Y	—	T	T	T	T	T
4	Y	Y	Y	—	T	T	T	T
5	Y	Y	Y	Y	—	T	T	T
6	Y	Y	Y	Y	Y	—	T	T

(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{27}{48} + \frac{8}{48} - \frac{5}{48} = \frac{30}{48}$.
(c) Not independent. $P(A) = \frac{27}{48} = 0.5625$ does not equal $P(A \mid B) = \frac{5}{8} = 0.625$.

T5.12 **(a)**



(b) $P(\text{defective}) = (0.60)(0.10) + (0.30)(0.30) + (0.10)(0.40) = 0.19$. **(c)** Machine B. $P(A \mid \text{defective}) = \frac{0.06}{0.19} = 0.3158$. $P(B \mid \text{defective}) = \frac{0.09}{0.19} = 0.4737$. $P(C \mid \text{defective}) = \frac{0.04}{0.19} = 0.2105$.

T5.13 **(a)** Here is a two-way table that summarizes this information:

	Smokes	Does not smoke	Total
Cancer	0.08	0.04	**0.12**
No cancer	0.17	0.71	**0.88**
Total	**0.25**	**0.75**	**1.00**

$P(\text{gets cancer} \mid \text{smoker}) = \frac{0.08}{0.25} = 0.32$.

(b) $P(\text{smokes} \cup \text{gets cancer}) = 0.25 + 0.12 - 0.08 = 0.29$.

(c) $P(\text{cancer}) = 0.12$, so $P(\text{at least one gets cancer}) = 1 - P(\text{neither gets cancer}) = 1 - 0.88^2 = 0.2256$

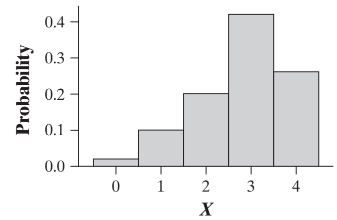
T5.14 **(a)** Let 00−16 represent out-of-state and 17−99 represent in-state. Read two-digit numbers until you have found two numbers between 00 and 16. Record how many 2-digit numbers you had to read. **(b)** The first sample is 41 **05 09** (it took three cars). The second sample is 20 31 **06** 44 90 50 59 59 88 43 18 80 53 **11** (it took 14 cars). The third sample is 58 44 69 94 86 85 79 67 **05** 81 18 45 **14** (it took 13 cars).

Chapter 6

Section 6.1

Answers to Check Your Understanding

page 350: **1.** $P(X \geq 3)$ is the probability that the student got either an A or a B. $P(X \geq 3) = 0.68$.
2. $P(X < 2) = 0.02 + 0.10 = 0.12$
3. The histogram below is skewed to the left. Higher grades are more likely, but there are a few lower grades.



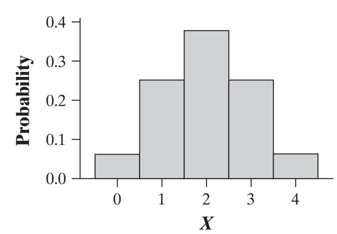
page 355: **1.** $\mu_X = 1.1$. If many, many Fridays are randomly selected, the average number of cars sold will be about 1.1. **2.** $\sigma_X = \sqrt{0.89} = 0.943$. The number of cars sold on a randomly selected Friday will typically vary from the mean (1.1) by about 0.943 cars.

Answers to Odd-Numbered Section 6.1 Exercises

6.1 **(a)**

Value	0	1	2	3	4
Probability	1/16	4/16	6/16	4/16	1/16

(b) The histogram below shows that this distribution is symmetric with a center at 2.



(c) $P(X \leq 3) = 15/16 = 0.9375$. There is a 0.9375 probability that you will get three or fewer heads in 4 tosses of a fair coin.
6.3 **(a)** $P(X \geq 1) = 0.9$. **(b)** The event $X \leq 2$ is "at most two nonword errors." $P(X \leq 2) = 0.6$. $P(X < 2) = 0.3$.
6.5 **(a)** All of the probabilities are between 0 and 1 and they sum to 1. **(b)** The histogram below is unimodal and skewed to the right.